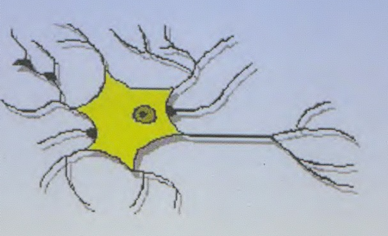
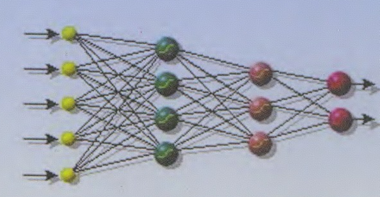


الشبكات العصبونية الصنعية

بين النظرية والتطبيق

الجزء الأول





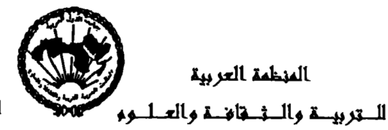
إعداد

الدكتور المهندس أحمد الكرمو

مراجعة

الدكتور المهندس راكان رزوق

الدكتور المهندس رامز حاج إسلام



المركز العربي
للتعريب والترجمة والتأليف والنشر

الشبكات العصبونية الصنعية

بين النظرية والتطبيق

الشبكات العصبونية الصنعية

بين النظرية والتطبيق

الجزء الأول

إعداد

الدكتور المهندس أحمد الكرمو

مراجعة

د.م. راكـــان رزوق **د.م. رامز حاج إسلام**

دمشق 2001

الشبكات العصبونية الصنعية بين النظرية والتطبيق - الجزء الأول
تأليف: د. أحمد الكرمو
المركز العربي للتعريب والترجمة والتأليف والنشر بدمشق
ص.ب: 3752 ـ دمشق ـ الجمهورية العربية السورية
هاتف: 3334876 11 963+ ـ فاكس: 3330998
E-mail: acatap@net.sy
Web Site: www.acatap.htmlplanet.com

المحتويات

الفصل الثاني: خواص الشبكات العصبونية الصنعية

الفصل الثالث: مراجعة عامة لبعض المفاهيم الرياضية والإحصائية

الفصل الرابع: أولى الشبكات العصبونية الصنعية وتطورها

الفصل الخامس: شبكات الذاكرة المترافقة

الفصل السادس: الشبكات العصبونية متعددة الطبقات بتغذية أمامية والانتشار الخلفي

بســـم اللـه الـرحمـن الرحيم

مـــقدمـــة

كُتب هذا العمل للقادمين الجدد إلى حقل الشبكات العصبونية الصنعية الذين ليست لديهم معرفة حول هذا الموضوع، إضافة إلى أولئك الذين اكتسبوا معرفة سابقة في هذا المجال. تتضمن الفصول الأولى للقادمين الجدد مواضيع تمهيدية عن عناصر الحساب البسيطة وخوارزمية التعليم الأساسية. وتتناول هذه الفصول أيضاً التطورات التاريخية التـــي مرت بها الأبحاث في تصميم وتطوير هذه الشبكات، وتُظهر الدواعي البيولوجية التـــي دفعت الباحثين للمضيِّ قدماً في مجال الشبكات العصبونية الصنعية. أما أولئك الذين لديهم خبرة سابقة في الموضوع، فإن هذا الكتاب يعتبر أرضاً خصبةً حيث يوفّر حقلاً واسعاً للاّعبين المهرة لإظهار قدراتهم في التعليم والتصميم والتطبيق؛ إذ يشتمل هذا الكتاب على دراسة نظرية وتطبيقات عملية لأكثر من ثلاثين نوعاً من بنـــى الشبكات العصبونية، ويرصد تطور البحث في هذا المجال منذ بزوغ فجره عام 1943.

ونود التنبيه إلى أن دارس هذا الكتاب يجب أن يكون لديه مستوى عالٍ لبعض المفاهيم الرياضية، وخاصة في الجبر الخطي، والمعادلات التفاضلية، والاحتمالات، ونظرية المعلومات، ونظرية الزمرة العائمة، والمنطق العائم، وقد آثرنا إضافة فصل خاص بتلك المفاهيم السابقة لتكون في متناول القارئ على نحوٍ مختصر يفي بالغرض.

أُقرُّ أن دوري في هذا العمل هو الترجمة لأحدث ما نشر في مجال الشبكات العصبونية الصنعية بعد دراسة هذا الموضوع وتدريسه لطلاب دبلوم الدراسات العليا في قسم هندسة الاتصالات–كلية الهندسة الكهربائية والإلكترونية–جامعة حلب، وقبل ذلك اطلاعي على هذا الموضوع وأهميته في شتـــى المجالات كالعلوم والهندسة والطب والتجارة والمال من خلال دراستـــي للدكتوراه بين أعوام 1992-1997 في المدرسة الوطنية العليا للإلكترونيات الصناعية والإلكترونيات الدقيقة والمعلوماتية وميكانيك السوائل التابعة للمعهد الوطنـــي للتكنولوجيا

بمدينة تولوز– فرنسا (ENSEEIHT-INPT).

كلنا أمل أن يكون هذا الكتاب مرجعاً في الشبكات العصبونية الصنعية، والحجر الأساس لوحدات البحث العلمي الأكاديمية في عالمنا العربي، حيث أثبتت الأبحاث في هذا المجال وخلال الخمسين سنة الماضية جدارة ومقدرة عالية لهذه الشبكات في مختلف مجالات التطبيقات العملية؛ كالاتصالات والهندسة والقانون وعلوم الحاسوب والتحكم والإحصاء والطب والصناعة والنقل والمالية و...الخ، حتى قيل إن تطبيقات الشبكات العصبونية الصنعية لا تعد ولا تحصى.

في النهاية لا يسعني إلى أن أتقدم ببطاقة تقدير وعرفان بالجميل إلى جميع الباحثين في مجال الشبكات العصبونية الصنعية الذين تعلمت منهم هذا الحقل من العلوم. كما أتقدم بالشكر الجزيل والامتنان إلى المركز العربي للتعريب والترجمة والتأليف والنشر ممثلاً بالسيد الدكتور غسان حلبوني لرعايته هذا العمل من ألفه إلى يائه.

وأتقدم بالشكر والعرفان بالجميل إلى السيد الدكتور راكان رزوق والسيد الدكتور رامز حاج اسلام اللذين يعود لهما الفضل الكبير في التدقيق العلمي لهذا العمل. أخيراً، أتقدم بالشكر لزوجتي التي احتملت على مضض هذا الضيف الثقيل الذي حل في بيتنا عاماً كاملاً.

والله من وراء القصد وهو ولي التوفيق

الدكتور المهندس أحمد الكرمو

جامعة حلب – كلية الهندسة الكهربائية والإلكترونية

الفصل الأول

مدخل إلى الشبكات العصبونية الصنعية

نقدم في هذا الفصل التمهيدي وصفاً للعمل الأساس لشبكة عصبونية صنعية بسيطة البنية، وأنواع الحسابات التـــي يمكن أن تنجزها. كما نناقش خواص الأجهزة العصبونية البيولوجية، وكيف ألهمت الباحثين في هذا المجال لبناء نماذج (models) الشبكات العصبونية، التـــي دعيت بالشبكات العصبونية الصنعية ANNs (Artificial Neural Networks).
سنتبع ذلك لمحةً تاريخية لنرى كيف تطور هذا الحقل من العلوم عبر الخمسين سنة الماضية.
أخيراً، سنقدم وصفاً لبعض التطبيقات العامة التـــي استخدمت فيها الشبكات العصبونية الصنعية بنجاح كبير، وسنعرض الأجزاء الأساسية والوظيفية لشبكة عصبونية بسيطة البنية وأنواع المهام التـــي تستطيع تعلمها وإنجازها.

1.1 تمهيد

الشبكات العصبونية الصنعية هي عبارة عن نماذج مبسطة للجهاز العصبـــي المركزي عند الإنسان. إنها شبكات بعناصر حساب عصبونية عالية الوصلات الداخلية فيما بينها لها المقدرة على الاستجابة لإشارة أو منبه الدخل (input stimuli) والتعلم لتتلاءم مع الوسط المحيط.
يعتقد الكثير من الباحثين في هذا المجال أن نماذج الشبكة العصبونية الصنعية تقدم آمالاً كبيرة تقربنا من بناء أنظمة حاسوب ذكي حقاً؛ حيث إن استعمال الحسابات التفرعية الموزعة كتلك المنفذة في الشبكات العصبونية الصنعية هي الطريقة الفضلى للتغلب على الانفجار التركيـــي (combinatorial explosion) المرافق للحسابات التسلسلية الرمزية في بنيان حاسوب فون نيومان (Von Neumann).
يوفر جهاز الشبكة العصبونية عند الإنسان أداة قوية لمصلحة هذا الموضوع، فالشبكات

البيولوجية قادرة على معالجة الملايين من إشارات (أو منبهات) الدخل في زمن من رتبة عشـــرات الميللي ثانية (10^{-3} ثانية)، مع أن هذه المعالجة ذات طبيعة إلكتروكيميائية (Electrochemical) وانتشارها يجري بمعدلات بطيئة نسبياً (من رتبة الميللي ثانية).

طبعاً هذه المعدلات أبطأ بمرات عديدة من العمليات السريعة جداً (من رتبة البيكو ثانية؛ 10^{-12} ثانية) في الحواسيب الرقمية التسلسلية التقليدية. رغم هذا الفرق الشاسع في معدلات انتشار الإشارة وسرعة معالجة الوحدة، فإن أنظمة الحاسوب التقليدي، كأنظمة الرؤية مثلاً، تبقى خلف الإنجاز الذي تقوم به الأنظمة البيولوجية في مقدرتها على المعالجة.

أثبتت الشبكات العصبونية الصنعية أنها معالجات حسابية فعالة في مسائل عديدة ضمن حقول متنوعة تشمل: تعرف الأشكال (pattern recognition)، وتعرّف الكلام والصور المرئية مثلاً، والتصنيف (classification)، والاستدعاء المترافق (associative recall)، وضغط المعطيات (data compression)، والتنبؤ (forecasting)، والنمذجة (modeling)، والتحكم المتكيف (adaptive control) وترشيح الضجيج (noise filtering)، وحل المشكلة التركيبية (combinatorial problem)، والطب (medicine)، ودمج معطيات عدة حساسات (multisensor data fusion)...الخ.

تقدم لنا الشبكات العصبونية الصنعية عدداً من الخواص المرغوب فيها غير الموجودة في أنظمة الحساب الرمزية التقليدية، فمن هذه الخواص، الإنجاز الموثوق عند التعامل مع نماذج دخل ضجيجية جزئية(غير كاملة)، ودرجة عالية للتسامح مع الخطأ (fault tolerance)، ومعدلات حساب تفرعية عالية السرعة، والمقدرة على التعميم (generalization)، وعلى التعلم المتكيف (adaptive learning).

سنصف في هذا الفصل التمهيدي نظاماً لشبكة عصبونية بسيطة لنوضح من خلاله كيف يعمل كمعالج تفرعي، وما هي المبادئ الأساسية لطرائق تعليم الشبكات العصبونية الصنعية.

وسنقدم ملخصاً عن خواص العصبونات البيولوجية، وكيف أوحت تلك العصبونات الحية بتطور شبكات المحاكاة (simulated networks)، بالإضافة إلى شبكات التكامل وسيع النطاق؛ شريحة VLSI (Very Large Scale Integration).

ولن ننسى أن نقدم لمحة تاريخية عامة عن الحوادث الهامة في أبحاث الشبكات العصبونية الصنعية التـــي انتقلت من مستوى البحث النظري إلى التطبيق العملي في البنـــى الحديثة هذه الأيام. وسنختم هذا الفصل بعرض مختصر لأهم تطبيقات هذه الشبكات (تقدم الفصول اللاحقة في هذا الكتاب تطبيقات كثيرة ومفصلة للشبكات ككل).

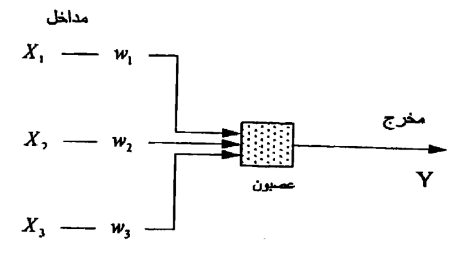
2.1 مفاهيم أساسية في الحساب العصبوني

على الرغم من أن بنـــى الشبكات العصبونية الصنعية يختلف بعضها عن بعض كثيراً فإن العصبون النموذجي أو عنصر الحساب (processing element)، وقد يسمى أيضاً عصبوناً أو وحدةً أو خليةً أو عقدةً، هو أساساً مقارِنٌ (comparator) يُعطي خرجاً عندما يتجاوز التأثير التراكمي (cumulative effect) لإشارات دخله قيمة عتبة معينة (threshold).

تتألف الشبكة العصبونية الصنعية، عموماً، من عدد كبير من عناصر المعالجة البسيطة المسماة بأحد الألفاظ الخمسة السابقة، وكل عصبون متصل مع عصبونات أخرى بواسطة خط اتصال موجه، وكل خط له وزن (weight) مرافق. يوضح الشكل (1.1) مخطط عصبون مفرد بثلاثة مداخل (X_1 , X_2 , X_3) وخرج وحيد هو Y. كل خط i، حيث $i = 1, 2, 3$، مغذى بإشارة دخل خارجية x_i وكل وصلة مرفقة بوزن موافق w_i؛ عبارة عن مرشح يعمل في دارة ربط الدخل مع العصبون.

يمكن أن تكون قيم الدخل حقيقية موجبةً أو سالبة، وقد تكون قيماً ثنائية (0, 1) أو قيماً ثنائية القطبية (1−, 1+).

تعمل الأوزان، التـــي تمثل وصلات العصبون الليفية في الشبكة العصبونية الحية، إما على زيادة (تهييج) (excitatory) أو إنقاص (تخميد) (inhibitory) إشارات دخل العصبون. يمكن أيضاً أن تكون قيم الأوزان حقيقية، أو ثنائية، ولكنها تفرض عادة حقيقية موجبة لخطوط التهييج وسالبة لخطوط التخميد. ويمكن أن تكون قيم خرج الشبكة حقيقية، أو ثنائية، أو ثنائية القطبية. يوضح (الشكل 2.1) مخططاً مكافئاً لعصبون وحيد.

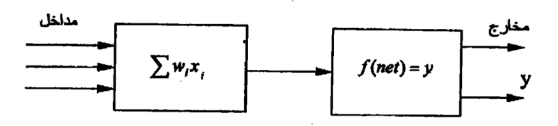


الشكل 1.1: شبكة عصبونية صنعية

حيث يعمل العصبون كتابع تنشيط (Activation) أو تابع تطبيق $f(.)$ (mapping) معطياً خرجاً يساوي:

$$y = f(net) \tag{1.1}$$

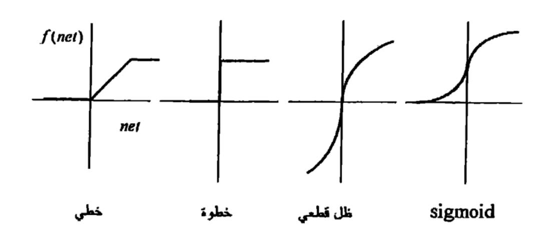
حيث *net* هو قيمة الدخل التراكمي للعصبون (يسمى عادة بدخل الشبكة)، ويكون *f* عادة تابعاً غير خطي يسمى بتابع التفعيل (activation function). فمثلاً، يؤخذ *net* غالباً كمجموع مُثقّل (weighted sum) لقيم المداخل كما يلي:



الشكل 2.1: مخطط تمثيلي لعصبون وحيد

$$net = x_1 w_1 + x_2 w_2 + x_3 w_3 = \sum_{i=1}^{3} x_i w_i \tag{2.1}$$

يمتاز *f* عادة بأنه متزايد ومستمر، ويوضح (الشكل 3.1) بعض الأمثلة عن توابع التفعيل $f(net)$ التـــي سنشرحها بالتفصيل لاحقاً.



الشكل 3.1: توابع تفعيل نموذجية لشبكة عصبونية

تتميز الشبكة العصبونية الصنعية بثلاثة تعابير أساسية: أولاً نموذج الوصل بين العصبونات، وهذا يسمى بالبنية (architecture)، وثانياً بطريقة تعيين الأوزان المرافقة للوصلات، وهذا يسمى بالتدريب (training) أو التعليم (learning) أو الخوارزمية (algorithm)، وثالثاً بتابع التفعيل (activation).

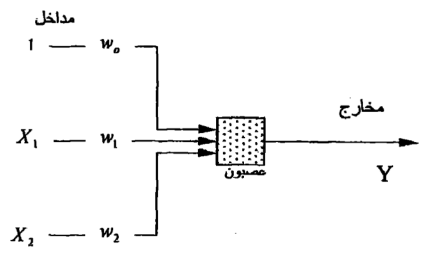
في التطبيقات العملية يكون عدد مداخل الشبكة أكبر بكثير من ثلاثة، لذا سنستعمل الدليل n للدلالة على عدد المداخل الكيفي.

لاحظ أن قيمة العتبة θ تكون أحياناً متضمنة في تعريف التابع *net* وذلك بوضع انحياز (bias) ثابت للدخل، وهو بقيمة +1 دائماً، على أحد خطوط الدخل، ووضع القيمة $w_0 = -\theta$. يوضح (الشكل 4.1) بنية شبكة عصبونية مع انحياز للدخل.

حيث w_0 قيمة الوزن المرافق لدخل الانحياز. يستطيع عصبون بسيط كالموضح في (الشكل 1.1) حساب العلاقة بين نوعين مختلفين فقط لأشكال الدخل وتصنيفهما ضمن مناطق منفصلة.

تعطى مركبات شعاع الدخل x بواسطة القيم $x_1, x_2, \ldots, x_n$، وهي عبارة عن قيم دخل منفصلة ومستقلة بعضها عن بعض، وتكون هذه القيم موافقة لحالات ونوعية الدخل. مثلاً، قد تكون x_i خواصٌ لرؤية الأشياء (شدة الإضاءة) في حقل الرؤية، أو خواص الإشارة الكلامية (طيف القدرة لترددات مختلفة)، أو قيماً فيزيائية كمتحولات عملية ما (درجة

الحرارة، الرطوبة، معدل التدفق، ...الخ)، أو القيم المحددة للوضع المالي لطالب قرض (عمره، دخله، الأقساط الشهرية، وضعه الوظيفي، تاريخ القرض، ...الخ)، أو القيم الفيزيائية الضرورية لتمييز حالة الطقس اليومي (وهكذا تستطيع الشبكة العصبونية التنبؤ بطقس الغد)، أو الأعراض المرضية المقيسة والمراقبة لشخص مريض (حرارته، ضربات قلبه، تحليل دمه، لون بوله، ...الخ).



الشكل 4.1: شبكة عصبونية مع انحياز للدخل

أما خرج الشبكة y فسيكون موافقاً لحالة الدخل المطبق؛ كاستمرارية الشيء المرئي، كلمة من نص، أو معدل التحكم بتدفق سائل، أو مستوى المجازفة (risk level) لطالب القرض، أو حالة الطقس المتنبأ به (صحو/ممطر)، أو الحالة الصحية لمريض (سليم/مريض).

يتصرف العصبون في حساب العلاقة بين شكلين مختلفين للدخل كمتنبئ، حيث يحسب مفهوم العلاقة أو فكرة العضوية بين أشكال الدخل.

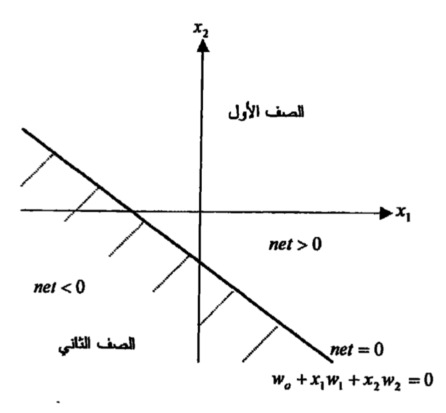
تُعرِّف قيمة شعاع الوزن **w** مستوى الفصل في الفراغ ذي البعد **n**، حيث يقسِّم قيمَ الدخل إلى منطقتين أو صفين (class-1) و(class-2) وفقاً لقيمة المجموع المُثقَّل $\mathbf{x}^T\mathbf{w} = \sum_i x_i w_i$ هل هي أكبر من الصفر أو أقل من الصفر (أو أية عتبة أخرى مثل θ):

إذا كان $\mathbf{x}^T\mathbf{w} = \sum_i x_i w_i \quad > 0$ فإن **x** ينتمي إلى الصف الأول

إذا كان $\mathbf{x}^T\mathbf{w} = \sum_i x_i w_i \quad \leq 0$ فإن **x** ينتمي إلى الصف الثاني

طبعاً سيكون تصنيف المناطق وفقاً لكل نوع من إشارات الدخل سهلاً جداً في علاقة

ثنائية البعد n = 2 كما هو موضح في (الشكل 5.1). تكون قيمة دخل الانحياز عادةً مساوية للواحد، ولكن فعالية هذا الدخل يُعبَّر عنها بقابلية تعديل وزن الانحياز w_0 كسائر الأوزان.



الشكل 5.1: مستوى الفصل لعلاقة ثنائية

إن وزن الانحياز مع أوزان المداخل الأخرى تحدد المحل الهندسي للخط الفاصل بين مناطق صفّي إشارات الدخل. يمكن أن يظهر هذا بسهولة عندما تكون معادلة الدخل ***net*** عبارة عن معادلة خط مستقيم في فراغ ثنائي البعد (x_1, x_2):

$$\text{net} = w_0 + x_1 w_1 + x_2 w_2 = 0$$

أو

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{w_0}{w_2} \qquad (3.1)$$

من هذه المعادلة نجد بسهولة ميل الخط الفاصل $-w_1/w_2$، وستكون إزاحة هذا الخط عن المبدأ محددة بواسطة وزن الانحياز w_0، وهكذا يكون المحل الهندسي وَجهة إزاحة الخط الفاصل، بين مناطق أنواع الدخل في الفراغ الثنائي البعد، قد تعيّن تعييناً كاملاً انطلاقاً من

قيم شعاع الوزن **w**.

إذاً ستكون قيم شعاع الدخل **x** لشعاع وزن ثابت **w** مصنفةً بانتمائها إما إلى الصف الأول، عندما تكون قيمة $net > 0$، أو إلى الصف الثانـــي عندما تكون قيمة $net \leq 0$.

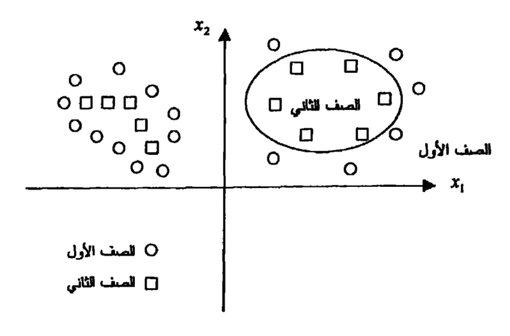
وهكذا نرى إن عملية اختيار مجموعة محددة من الأوزان يعين الحدود الفاصلة بين أشكال الدخل **x** المختلفة، وهذه هي الأهمية الكبيرة التـــي تعتبر مفتاح التعليم الذاتـــي في الشبكات العصبونية الصنعية.

أي أن المعرفة الضرورية لإجراء عملية التصنيف بين أشكال الدخل تكمن في قيم الأوزان. يمكن إيجاد الحدود الفاصلة بين مجموعة من أشكال الدخل بتعديل قيم شعاع الأوزان **w** وفقاً لمعايير التصنيف المرغوب بها. في الفصول اللاحقة سنتعلم طرائقاً مختلفة تستطيع فيها الشبكة العصبونية الصنعية تعلم التطبيق **(mapping)** لحدود الفصل بين الصفوف بواسطة عملية تعديل للأوزان.

إن الفواصل الخطية المستعملة للفصل بين المناطق أو الصفوف، التـــي تجزئ فراغ الدخل إلى جزأين أو أكثر، هي عبارة عن خطوط أو مستويات خطية تعطى بواسطة الجداء $\mathbf{x}^T\mathbf{w}$ (يرمز T إلى مرافق الشعاع).

بوجه عام، لا يمكن فصل المناطق المحاطة بمنحنيات متداخلة بعضها مع بعض بفواصل خطية، وتفصل مثل هذه المناطق بتوابع غير خطية كما هو موضح في (الشكل 6.1).

إن كل عنصر خرج من شبكة عصبونية خطية بتغذية أمامية من النوع المشروح سابقاً (طبقة واحدة من عناصر الدخل المتصلة بواسطة الأوزان إلى طبقة الخرج الوحيدة) يستطيع فصل الفراغ إلى صفوف مفصولة خطياً فقط، على حين يمكن النظر إلى المسائل اللاخطية على أنها تركيب لمسائل خطية، ومن ثَم فلحلّ مثل هذه المسائل يمكن إضافة طبقة أو أكثر (بين الدخل والخرج) أو استعمال توابع تركيب غير خطية إما في نفس الطبقة وإما في الطبقات المتتالية المضافة. سنوضح ذلك في بنـــى الشبكات العصبونية الصنعية الأكثر تعقيداً القادرة على التغلب على كل المصاعب المذكورة آنفاً.



الشكل 6.1: مناطق صف ليست مفصولة خطياً

3.1 الشبكات العصبونية البيولوجية

لقد استلهمت الأبحاث في مجال الشبكات العصبونية الصنعية وتأثرت بمعرفتنا عن الأجهزة العصبونية البيولوجية عند الكائنات الحية. بالطبع لا تزال هذه المعرفة بالجهاز العصبي محدودة جداً، إلا أن علماء الجهاز العصبي اكتشفوا بعض الحقائق الهامة خاصة خلال العقود القليلة الماضية.

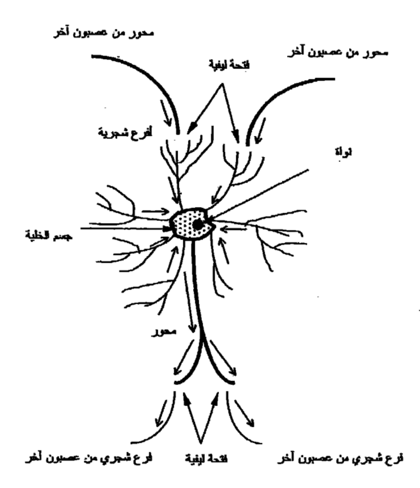
يعتبر العصبون البيولوجي عنصر الحساب الأساسي في الأنظمة البيولوجية الحية، فما هو العصبون؟ وكيف يعمل؟ العصبون هو خلية صغيرة جداً تستقبل تنبيهاً إلكتروكيميائياً من منابع عديدة، وتستجيب بتوليد نبضات كهربائية ترسل إلى عصبونات أخرى أو خلايا مؤثرة (effector cells).

هناك ما يقرب من 10^{+10} إلى 10^{+12} عصبون في الجهاز العصبي لدى الإنسان، يستطيع كل منها تخزين عدة بتات (bits) من المعلومات. ولما كان متوسط الوزن الكلي للدماغ قرابة 1.5 كغ، فإن متوسط وزن العصبون أقل من 1.5×10^{-9} غرام.

تستقبل العصبونات إشارات الدخل (أو التنبيه) من الخلايا الحسية أو من خلايا أنواع أخرى، وترسل المخارج إلى عصبونات أخرى أو أعضاء مؤثرة كالعضلات أو الأطراف... إلخ.

إن قرابة 10% من العصبونات هي عصبونات دخل (مُورِّد) (afferent) وخرج (مُصدِّر) (efferent) والـ 90% الباقية هي وصلات داخلية مع العصبونات الأخرى التـــي تخزن المعلومات أو تؤدي إلى تحويلات متنوعة على الإشارات المرسلة خلال الجملة العصبية. على الرغم من الأنواع المختلفة العديدة للعصبونات التـــي شُخِّصت وعرفت، فإنها تشترك بخواص عامة، كما هو موضح في (الشكل 7.1).

فالعصبونات خلايا معقدة تستجيب لإشارات إلكتروكيميائية. يتألف العصبون من نواة (Nucleus)، وجسم الخلية (Soma)، وخطوط عديدة متفرعة شجرياً (Dendrites) تعتبر وصلات دخل من عصبونات أخرى عبر الألياف (Synapses)، ومحور (Axon) يحمل كمون الخرج الفعال إلى عصبونات أخرى عبر خطوط النهايات والألياف، كما هو موضح في (الشكل 7.1).



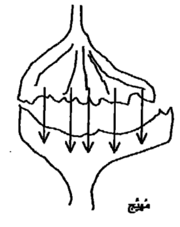
الشكل **7.1**: بنية عصبون نموذجي مع اتجاه إرسال الإشارة بين العصبونات

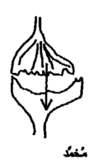
إن الفعالية العصبونية متعلقة بإيجاد كمون كهربائي داخلي يدعى"كمون الغشاء" (membrance potential)، ويمكن أن يزداد هذا الكمون أو ينقص وفقاً لفعالية الدخل المستقبل من خلايا أخرى عبر الألياف.

إذا ازدادت قيمة المداخل التراكمية حتـــى يصل الكمون إلى قيمة العتبة عندها يتنشط العصبون، فيطلق سلسلة من النبضات الشوكية الكمونية إلى أسفل المحور لتهييج أو تخميد عصبونات أخرى. يبلغ معدل بث النبضات حوالي 10-125 ميللي ثانية ويبلغ الزمن اللازم لكي يَعبر التنبيهُ الليفَ حوالي 1 ميللي ثانية.

تجري الوصلات عبر نوعين من الألياف هما:

المُهيِّج (excitatory)، والمُخمِّد (inhibitory)، كما هو موضح في (الشكل 8.1).





الشكل 8.1: نوعان عامّان للألياف

بعد تنشيط العصبون هناك فترة استرخاء تبلغ حوالي 10 ميللي ثانية لا يستطيع خلالها التنشط ثانية. تقاس فعالية العصبون بتردد النبضات الشوكية الكمونية التـــي يولدها، وهي تقع بين 50 حتـــى بضع مئات من النبضات خلال الثانية الواحدة.

رغم معرفتنا المتواضعة عن عمليات التعليم ضمن الدماغ، إلا أنه من المعتقد أن نموٌّ تفاعل حيوي ما يحدث في العصبونات نتيجةً لازدياد فعالية الخلية، ومن المعتقد أيضاً أن هذا النمو هو المسؤول عن الذاكرة والتعليم داخل الدماغ. فهو يؤثر في شحنة الكمون ليعطي الليف ما يعتقد أنه المكافئ للأوزان المستعملة في الشبكات العصبونية الصنعية.

قد تكون مساحة الليف مرتبطة بالتعليم، فهي تتغير بواسطة عامل ضرب يزيد على عشرة. ومن المعتقد أن الألياف ذات المساحات الكبيرة تكون مُهيَّجة، على حين تكون الألياف ذات المساحات الصغيرة مُخمَّدة كما هو موضح في (الشكل 8.1).

كان Hebb أول من اقترح عام 1949[1] أن مثل هذه العملية مرتبطة بالتعليم في الأجهزة البيولوجية، وقد أصبح تعليل Hebb عن كيفية تحقق النمو ملهماً لعدة قواعد تعليم مختلفة للشبكات العصبونية الصنعية عبر السنين.

سنلخص الآن بعض خواص الشـــبكات العصبونية البيولوجية عند الإنسان في الجدول التالي :

عدد العصبونات	10^{+13}-10^{+11}
عدد الوصلات	مئات – مئات الآلاف
عدد المداخل	%10
عدد المخارج	%90
سعة التخزين	(10^{+15}-10^{+13}) خانة
عامل الاستعمال	%10
متوسط وزن الدماغ	1.5كغ
متوسط وزن العصبون	1.2×10^{-9} غرام
معدل انتشار الإشارة	125-5 م/ثانية
زمن عبور الليف	1 ميللي ثانية
دور الاسترخاء	10^{-2} ثانية
تردد التنشيط	100-50 نبضة/ثانية
الأنواع العامة للألياف	مُهيَّج ومُخمَّد
كمون الغشاء	قدحات التنشيط
تغير مساحة الليف	بعامل عشرة
حالة العمل	غير متزامن

4.1 تاريخ البحث العلمي في الشبكات العصبونية الصنعية

بدأ التاريخ الحديث للشبكات العصبونية الصنعية عام 1943، عندما نشر McCulloch وزميله Pitts مقالتهما الهامة التـــي وَصفت عصبوناً بسيطاً ثنائي الحالة، (ثنائي مع عتبة)، فيه كلّا نوعي المداخل المهيِّجة والمخمِّدة.

استطاعت هذه الوحدات إنجاز التوابع المنطقية الأساسية مثل AND وOR وNOT...الخ. واعتُبِرت الشبكات بهذه الوحدات آنذاك نماذج تمثيلية للدماغ عند الإنسان.

كان Donald Hebb أول من اقترح قاعدة معقولة لطريقة تعليم العصبون، وذلك عام 1949. وحتـــى اليوم ما يزال الكثير من نماذج التعليم التـــي يستعملها الباحثون أمثلة لتطبيق قاعدة Hebb.

يجب ألاّ ننسى بأن Hebb هو أول من علل أن تجمعات خلوية موزعة عديدة استعملت لتمثيل المعرفة، وكان هذا من الاقتراحات الأولى المبكرة عن البنية التوصيلية للعصبونات (connectionist architecture).

من المحتمل أن أول محاكاة (simulation) بين الشبكات العصبونية الصنعية والحاسوب نشرها Rochester عام 1956 في مؤتمر Dartmouth، حيث يعتبر ذاك التاريخ البداية الرسمية للذكاء الصنعي AI (Artificial Intelligence).

إن أول من بحث في الإدراك (perception) هو Rosenblat من عام 1958 وحتـــى 1962. والشبكة الأساسية للـــمدرك أو المفسر (perceptron) هي عبارة عن وحدة منطقية مع عتبة، مركبة من ثلاث طبقات: طبقة دخل حساسة للضوء متصلة عشوائياً مع طبقة مرافقة (association) تتصل بدورها مع طبقة التصنيف (classification) أو طبقة استجابة الخرج. تَستعمل قاعدة تعليم المفسر تعديل الأوزان المتكرر، وهذا أقوى بكثير من قاعدة Hebb.

نشر Minsky وPapert عام 1969 كتاباً في الإمكانيات الحسابية لشبكة المفسر، وكان هذا الكتاب تحليلاً رياضياً أنيقاً أظهر التوابع المنطقية التـــي تستطيع الشبكة حسابها؛ وهي التوابع الخطية فقط (مثل AND و OR و...الخ) وتلك التـــي تعجز عن حسابها كالتوابع غير الخطية (مثل XOR). وكان هذا الكتاب كالعاصفة التـــي ذهبت بالموارد المالية المخصصة

للبحث العلمي في الشبكات العصبونية الصنعية أدراج الرياح و لم يبق إلا القليل من الباحثين الذين استمروا في أعمالهم أمثال Anderson وKohonen وGrossberg وAmari وwidrow وغيرهم.

كان Widrow أول من طور التطبيقات العملية للشبكات العصبونية، حيث صمم عصبوناً بسيطاً مشابهاً للمفسر أسماه أدلين (ADALINE)، وركّب عدة شبكات من (ADALINE) أسماها مادلين (MADALINE). في الوقت الحالي هناك تطبيقات هامة لمثل هذه الشبكات مثل مضعفات الصدى المتكيفة للدارات الهاتفية العابرة للقارات، ومرشحات الضجيج في أجهزة المودم (تعديل/فك تعديل) (MODEMs) العالية السرعة، وأنظمة الهوائي المتكيف لهدف دوار غير متغير (عام 1967)، ولمسائل متنوعة في التحكم، كتوازن مكنسة وتراجع شاحنة (1987-1989).

خلال السبعينيات قام عدد من الباحثين وخاصة بعد نشر كتاب Minsky وPapert بالبحث في الشبكات العصبونية الصنعية كذاكرة مترافقة أمثال Kohonen وAnderson عام 1972. وكانت النماذج المستعملة في هذه الأبحاث خطية وأحادية الطبقة.

طوّر Anderson عام 1977 ذاكرة مترافقة قابلة للعنونة بالمحتوى content-addressable associative memory. وقد اعتمدت كل هذه النماذج من الشبكات على قاعدة تعليم Hebb.

ثم اكتشف Grossberg عام 1983 شبكات الطنين المتكيف.

يعتبر اكتشاف خوارزمية التعليم بتعديل الأوزان في شبكة التغذية العكسية متعددة الطبقات (وتسمى أحياناً المفسر متعدد الطبقات) واحداً من أهم التطورات في أبحاث الشبكات العصبونية الصنعية.

تُعرف هذه الخوارزمية بالانتشار الخلفي Bp (Backpropagation)، حيث تعدل الأوزان من طبقة الخرج باتجاه الوراء طبقة بعد طبقة حتـــى الوصول إلى طبقة الدخل، وذلك لتقليل الأخطاء في الخرج.

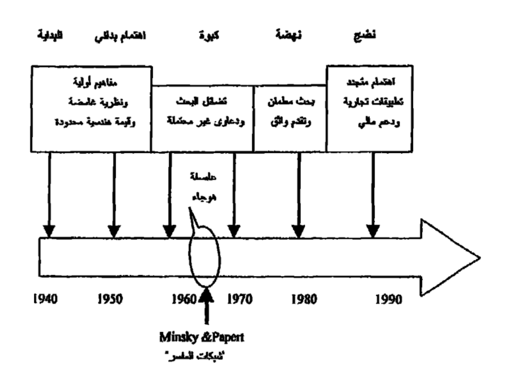
اكتُشفت هذه الطريقة خلال فترات زمنية مختلفة من قبل العديد من الباحثين أمثال Werbos عام 1974، وParker عام 1985، وRumelhart عام 1986. وفتحت خوارزمية

الانتشار الخلفي الطريق أمام حسابات عامة أكثر للشبكات العصبونية الصنعية، وذلك بواسطة التغلب على المصاعب التـــي كانت تعيق شبكات المفسر وأخواتها ذات الطبقة الوحيدة، وأصبحت هذه الشبكات قادرة على التعلم لحل المسائل غير الخطية مثل التابع XOR.

اخترع الباحث الياباني Fukushima شبكات (cognitron) و(neocognitron) عام 1988.

ومازالت الأبحاث في مجال الشبكات العصبونية الصنعية جارية على قدم وساق تتقدم بخطى واثقة وسريعة في التطبيقات العملية وذلك في شتـــى مجالات الحياة.

يلخص (الشكل 9.1) مجمل هذه الحوادث التاريخية.



الشكل 9.1: ملخص تطورات البحث العلمي في الشبكات العصبونية

5.1 تطبيقات الشبكات العصبونية الصنعية

سنعرض في هذا الفصل بعض التطبيقات العامة للشبكات العصبونية الصنعية لنعطي القارئ فكرة عن تنوع واتساع المجالات التـــي استعملت فيها هذه الشبكات. بالطبع سيكون عرضنا مختصراً، وعاماً، وسنقوم بمزيد من الشرح والتفصيل عند دراسة كل بنية على حده.

1.5.1 تحقيق الشروط المقيدة

تتطلب العديد من المسائل أن تكون الوسطاء المُعرَّفة للنظام مُقيَّدة ومُعيَّنة. فمثلاً، تحتاج عملية تصنيع ألواح الدارات الإلكترونية المطبوعة إلى ثقب أعداد ضخمة من الثقوب ، توضعة بدقة في الألواح. وسيكون الزمن الكلي اللازم لإتمام عملية الثقب تابعاً للنظام المتبع تثقيب، لأن هذا يحدد المسافة الكلية المقطوعة خلال عملية التثقيب. هذه القضية مشابهة شكلة البائع الجوال TSP (Traveling Saleman Problem)، التـــي تعتبر مسألة استمثال (Optimization)، وتحقيق شروط محددة. ستكون الشروط المقيدة بحيث أن كل ثقب يجب ن يزار مرة واحدة فقط، في حين يتحقق الاستمثال بتخفيض الزمن الكلي المصروف في ملية التثقيب؛ إذ يجب أن يكون زمن تثقيب اللوح منخفضاً لنحصل على حل أمثلي بهدف قليل الكلفة.

2.5.1 الذواكر القابلة للعنونة بالمحتوى Content-Addressable Memories

يمكن أن تعمل بعض بنـــى الشبكات العصبونية الصنعية كذواكر لتخزين أشكال لدخل، حيث يجري استرداد الشكل المخزن متـــى قدمنا للشبكة الشكل المصاحب. يمكن ن يكون للشكل المسترد نفس الشكل المخزن (ربما بنفس التشويه؛ أو بضجيج مضاف أو خانات محذوفة) أو شكلاً مختلفاً آخر.

نقول عن عملية الاسترداد إنها عملية استرداد مرافق ذاتـــي (Autoassociative retrieval) عندما تكون الأشكال المخزنة والمستردة واحدة، أما عندما تكون الأشكال لمخزنة والمستردة مختلفة، فإن عملية الاسترداد تسمى استرداداً مرافقاً مغايراً (Heteroassociative retrieval).

تسمى مثل هذه الشبكات ذواكر قابلة للعنونة بالمحتوى أو بمحتواها باعتبار أنها يمكن أن نستعمل لاسترداد الأشكال بواسطة أداة مقودة من محتوى النماذج المخزنة.

3.5.1 معالجة الإشارة

في حقل معالجة الإشارة هناك تطبيقات متعددة للشبكات العصبونية الصنعية. أحدها (وهو من أهم التطبيقات التجارية وما يزال) هو ترشيح الضجيج على الخطوط الهاتفية،

والشبكة المستخدمة لتحقيق هذا الغرض هي شكل من (ADALINE).

أصبحت الحاجة إلى مضعفات الصدى المتكيفة ضرورة ملحة، وذلك بعد تطور عالم الاتصالات وتوسع شبكات الاتصال الهاتفية القارية. وقد استُخدمت فكرة بسيطة جداً للحذف المتكيف للضجيج؛ حيث عند نهاية خط الهاتف للمسافات الطويلة تطبق الإشارة المستقبلة على كلا عنصري النظام الهاتفي، المسمى بالدارة الهجينة (hybrid)، والمرشح المتكيف (Adaptive Filter)، وهو عبارة عن شبكة (ADALINE). والفرق بين خرج الدارة الهجينة وخرج شبكة (ADALINE) هو الخطأ الذي سيستعمل لتعديل أوزان شبكة (ADALINE) وذلك لتقليل مقدار هذا الخطأ.

4.5.1 ضغط المعطيات

هناك تطبيق آخر في حقل الاتصالات، وهو استعمال بعض الشبكات العصبونية لتعلم التطبيق (mapping)؛ هذا يعنـــي إنقاص عدد أبعاد فراغ أشكال الدخل، ومن ثم إنجاز نوع من ضغط المعطيات.

تحوَّل الأشكال من فراغ الدخل ذي البعد n إلى فراغ ذي البعد m، حيث $m < n$، وذلك بواسطة تخصيص رموز codes لمجموعات الأشكال المتشابهة. تعمل كلمات الرمز في البعد m كنماذج أولية (prototypical) لكل تجمعات (clusters) أو مجموعات الأشكال المتشابهة في الفراغ ذي البعد n. ومن ثَم، يمكن أن نعالج أشكالاً بطول أقصر بكثير، وبذلك نكون قد أنقصنا عرض حزمة كمية المعلومات الواجب إرسالها في تطبيقات بث المعطيات، وأنقصنا سعة ذاكرة التخزين اللازمة عندما نخزن مجموعات الأشكال.

تظهر أهمية ضغط المعطيات، خصوصاً، في التطبيقات التـــي تجمع وتعالج كمية ضخمة من المعطيات، كما في حالة معالجة معطيات الصور الملتقطة عبر الأقمار الصناعية (معالجة الصور).

5.5.1 التشخيص

التشخيص هو تطبيق عام للشبكات العصبونية في حقول عديدة، كالطب والهندسة والصناعة..الخ. وتعتبر هذه القضية إحدى المهام الأساسية للتصنيف؛ حيث تتطلب ربطاً

صحيحاً بين أشكال الدخل، التـــي تمثل شكلاً ما من الأعراض أو السلوك غير الطبيعي، والمرض الموافق أو العطل الجهازي أو نوع آخر من التقصير الوظيفي. وتعتبر أنظمة التشخيص المعقدة، مثل أنظمة تشخيص الأمراض، من التطبيقات الشائعة للنظم الخبيرة.

بالطبع هناك تطبيقات عديدة متنوعة وحيوية باستعمال الشبكات العصبونية الصنعية منشورة في المقالات والكتب العلمية. والفكرة وراء هذه التطبيقات هي تدريب الشبكة للعمل كذاكرة ترافق ذاتي (تدعى شبكة حالة – دماغ – في صندوق، ستوصف لاحقاً بالتفصيل) لتخزين عددٍ ضخم من البطاقات الطبية، تتضمن كل بطاقة منها معلومات عن الأعراض والتشخيص والمعالجة النفسية أو الدوائية .

بعد التدريب، تغذى الشبكة بمجموعة من الأعراض؛ عندها ستعطي الشبكة على الخرج النموذج الكامل المخزن المرافق لتلك الأعراض المتضمن التشخيص الأفضل والدواء الأنجع .

يحقق مثل هذا التطبيق قضية العَولمة في معالجة الأمراض عبر الكرة الأرضية، وذلك في حال توفر بنك المعلومات (شبكة عصبونية كذاكرة ترافق ذاتـــي تخزن جميع الأمراض) المرتبط مع شبكة (Internet) والواصل إلى عيادة كل طبيب.

6.5.1 التحكم

استعملت الشبكات العصبونية الصنعية بفعالية في التحكم في الربوت المتجول في الهواء الطلق، بما في ذلك مهام قيادة المركبات بدون سائق (ذاتية القيادة).

وقد دربت هذه الشبكات على مهام صعبة مثل تفريغ عربة مقطورة في ميناء بأقل جهد، حتـــى عربات بمقطورتين مرتبطتين إحداهما مع الأخرى.

واستخدمت أيضاً للتحكم بنجاح في توضع الألكترودات الكبيرة في عمليات لحام القوس الكهربائي المستعملة في شركات تصنيع الفولاذ، حيث وفرت للشركات الملايين من الدولارات من خلال توفير الطاقة الكهربائية المستهلكة وإطالة عمر التجهيزات ذات الكلفة المالية العالية جداً.

تستعمل الآن الشبكات العصبونية الصنعية من أجل تحكم فعال ومؤثر في عملية إنتاج أطقم المنتجات الاستهلاكية (مجموعة من 12 أو 24 قطعة) وخاصة التـــي تصنعها الشركات

اليابانية. يبدو أن مجال تطبيقات الشبكات العصبونية الصنعية في حقل التحكم غير محدود!

7.5.1 التنبؤ

إن ظاهرة التنبؤ مهمة شائعة في مجالات عديدة: فمثلاً، تريد شركة منتجات استهلاكية معرفة النمو في مبيعاتها من أجل إنتاج جديد تريد طرحه في الأسواق، ويريد علماء الطقس توقعاً صحيحاً عن الطقس، وتريد البنوك التنبؤ عن الاعتمادات غير المستحقة للشركات لمنحها قروضاً، وتريد مجموعات إدارة المطارات معرفة نمو زبائنها الواصلين إلى المطارات المشغولة بالمسافرين، وتريد شركات الكهرباء معرفة مقدار تزايد طلبات مشتركيها من الطاقة الكهربائية مستقبلاً، وهكذا.

لقد أظهرت الشبكات العصبونية الصنعية نجاحاً فائقاً بوصفها أداة تنبؤ في مختلف المجالات؛ كالتنبؤ بأن حادثاً ما سيحدث أو لن يحدث، والتنبؤ في زمن وقوع حادث ما، والتنبؤ بمستوى حصيلة وقوع حادث ما.

للتنبؤ بمستوى مقبول من الدقة، يجب أن تدرَّب الشبكة العصبونية على مجموعات كبيرة من الأمثلة كالأزواج: حصيلة الشكل السابقة/اللاحقة مثلاً. ويجب أن تكون الشبكات العصبونية الصنعية فيما بعد قادرة على التعميم والاستقراء من الأشكال الجديدة التي ستقدم لها عن طريق الدخل للتنبؤ بنتائج وتأثيرات مرافقة.

يستعمل الكثير من المعاهد المالية والتجارية الآن الشبكات العصبونية الصنعية استعمالاً واسعاً في تجارة الأسهم والتبادل الأجنبي.

8.5.1 التطبيق العام (التخطيط)

إحدى الخواص البارزة لبعض الشبكات العصبونية الصنعية هي مقدرتها على تعلم توابع كيفية من مجموعة أمثلة التدريب، وتعطي هذه المقدرة مجالاً متنوعاً من التطبيقات. يمكن الإثبات بأن كل تطبيقات الشبكات العصبونية الصنعية تقع تحت معالجة عامة للتطبيق التابعي، حيث تتعلم الشبكات تحويل (تطبيق) شعاع دخل ذي n بعداً إلى شعاع ذي m بعداً وفقاً لبعض المعايير (معرفتها ليست ضرورية).

9.5.1 دمج معطيات عدة حساسات

إن عملية دمج معطيات لاقط أو حساس هي عملية جمع معطيات من منابع متعددة لكي نستخلص معلومات أكثر من خلال عملية الجمع هذه التـــي لا يمكن استخلاصها من المنابع منفردة.

تشمل عمليات الدمج للمعطيات الكشف (Detection)، والاقتران (Association)، والترابط (Correlation)، والتقدير (Estimation)، وضم المعطيات لإنجاز تقدير متطابق، وتخمين زمنـــي للأوضاع أو الحالات.

لقد تطورت تكنولوجيا الحساسات خلال الثمانينيات، وهذا ما أدى إلى انتشار واسع لتطبيقات دمج عدة حساسات. يشمل ذلك المعدات العسكرية، والتحكم في العمليات، والمراقبة، والربوت، والتشخيص، ومجالات أخرى.

المثال الأقوى في مجال دمج معطيات حساسات عديدة هو الإنسان وجهازه البيولوجي. يطبق الإنسان عملية دمج لمعطيات حواسه الجسمية (اللمس والرؤية والسمع والذوق والشم) ليكتسب إدراكاً يؤدي إلى تصرف له معنـــى في الوسط المحيط. تُجمع مئات الآلاف من اللواقط المعطيات في الزمن الحقيقي لدمجها ومعالجتها خلال مستويات استخراج تتابعية عالية. وينفّذ الجهازُ العصبـــي المركزي عمليةَ الدمج لمعطيات الحساسات بواسطة شبكة ضخمة من العصبونات المتصلة داخلياً بكثافة فيما بينها.

تقدم الشبكات العصبونية آمالاً كبيرة في تطبيقات دمج المعطيات، وذلك لنفس الأسباب لدى الأجهزة البيولوجية، حيث أثبتت جدارتها في مثل هذه المهام. لذا جعلت الشبكات العصبونية، ببنياتها المتنوعة ومقدرتها على المعالجة، الاختيار الطبيعي لتطبيقات دمج عديدة.

تنجز كل شبكة عصبونية صنعية نوعاً من دمج المعطيات للقيام بتخطيط مرغوب به لإشارات الدخل والخرج.

10.5.1 الاستمثال

استُعملت الشبكات العصبونية الصنعية في الكثير من المسائل التـــي تتطلب حلاً أمثلياً أو قريباً جداً من الأمثلي. تتطلب مثل هذه القضايا تحقيق بعض الشروط، وعلى الأغلب،

ستتداخل عملية البحث عن حل أمثلي مع بعض تطبيقات تحقيق الشروط المقيدة المشروحة سابقاً.

من تطبيقات الأمثلية: سعر وكمية بيع مقاعد المسافرين عبر الخطوط الجوية، وجدولة عمليات التصنيع (سلسلة من المهام للآلات لمواجهة بعض المعايير)، وإيجاد الطريق الأقصر عبر كل الرحلات الممكنة (مسافة الطريق) بين عدد كبير من مدن العالم أو النقاط الجغرافية الأخرى التـــي يجب أن تعبر بزيارة واحدة لكل رحلة، وتقليل تابع الكلفة تحت مجموعة من الشروط المحددة، وهكذا.

ومن الأمثلة القديمة التـــي استُخدمت فيها الشبكات العصبونية الصنعية، (مثل شبكة التكرار الديناميكي في حل مسألة الأمثلية) مسألة البائع الجوال من قبل Hopfield & Tank عام 1985[2].

11.5.1 تعرف الأشكال

تعتبر الشبكات العصبونية الصنعية وسيلة جيدة لتعلم النوع الإدراكي للمهام، مثل تعرف الأشكال والنماذج المختلفة المعقدة: كصورة مرئية للأشياء والأحرف المطبوعة أو المكتوبة باليد، وتعرف الكلام المنطوق من قبل أناس مختلفين، وأنواع أخرى في هذا المجال.

وما يزال الباحثون والمصممون يسعون لإيجاد تطبيقات ناجحة للشبكات العصبونية الصنعية في مجالات: معالجة الصور، وتعرف الكلام والكتابة اليدوية، وتعرف الأهداف آلياً، وفي التحكم والربوت، وهكذا.

وقد أثبتت عدة دراسات مقارنة، نفّذها كثير من الباحثين، أن الشبكات العصبونية الصنعية تعتبر مقنعة ومرضية إذا ما قورنت مع الطرائق الإحصائية التقليدية المتبعة في تعرف الأشكال.

الفصل الثانـــي

خواص الشبكات العصبونية الصنعية

وصفنا في الفصل السابق شبكة عصبونية صنعية بسيطة ذات ثلاثة مداخل وخرج وحيد. سنبحث في هذا الفصل البنية الصندوقية الأساسية للشبكات العصبونية الصنعية، وسندرس بعض مميزاتها الهامة وتطبيقاتها العامة وأيضاً مصطلحاتها العلمية. وسنناقش أيضاً بعض الخواص الهامة التـــي تميزت بها هذه الشبكات، وسنقدم لمحة عن طرق التعليم المختلفة المستخدمة في تدريب البنـــى المختلفة والمتنوعة للشبكات العصبونية الصنعية.

1.2 تمهيد

عرضنا في الفصل الأول واحدة من أبسط بنـــى الشبكات العصبونية الصنعية، وهي شبكة تغذية أمامية بطبقة واحدة؛ فهي أحادية الطبقة (single layer) لأن أشكال الدخل قد عولجت من خلال طبقة واحدة من العصبونات فقط (أو بالأحرى طبقة واحدة من الوصلات)، حيث تنتشر إشارة الدخل عبر وصلات مُثَقَّلة إلى العصبون فتتولد الاستجابة كتفعيل للخرج، وهي ذات تغذية أمامية (Feedforward) لأن الإشارات تنتشر فقط في الاتجاه الأمامي من عقد الدخل إلى عقد الخرج ولا يُسمح للإشارات بالانتشار الخلفي عبر العصبونات.

من الملاحظ أن مثل هذه الشبكات يمكن أن تعمل بأسلوب متزامن أو غير متزامن. ففي حالة العمل المتزامن، تتطور وسطاء الشبكة مع أمر خاص أو أزمنة خاصة (وهذا يعنـــي أن الكل لحظي). أما في حالة العمل غير المتزامن، فإن تطور الوسطاء للشبكة يكون عشوائياً، ويمكن التحكم في الزمن بين نقاط التطوير أو التحديث بواسطة توزيع احتمالي.

وهكـــذا نـــرى أن الوصف التام لسلوك الشبكة العصبونية الصنعية يتطلب إعطاء عدد

العصـــبونات، وكيفية الوصل الداخلي فيما بينها، والطريقة التـــي تغذى فيها الإشارات إلى مداخـــل العصبون، وتوابع التفعيل أو التطبيق لكل عصبون التـــي تنفذ على مداخله، والأمر الذي تنفذ فيه الحسابات، وأخيراً الطريقة التـــي تنتشر فيها الإشارات من الدخل إلى الخرج.

أخيراً سنحدد نظام المعادلات الديناميكية الضابطة لسلوك النظام وطريقة التعليم المستعملة لتدريب الشبكة.

2.2 بنـــى الشبكات العصبونية الصنعية
Artificial Neural Network Structures

لمعرفة خواص شبكة معطاة، من الضروري تعيين عدد العصبونات، وكيفية تحقيق الاتصال الداخلي فيما بينها، والمعالجة التـــي ستنفذ خلال الشبكة. ولما كانت هناك إمكانيات متنوعة وكثيرة في طريقة تركيب الإشارات ومعالجتها في الشبكة، فإننا سنقصر مناقشتنا فيما يلي على الطرائق الأكثر شيوعاً واستعمالاً في هذه الأيام، ومع ذلك سنرى أن هناك حوالي 20 إلى 30 نوعاً شائعاً للشبكات العصبونية الصنعية دُرست واستُعملت في تطبيقات مختلفة.

1.2.2 عناصر أو خلايا الحساب العصبوني
Neural computing elements or cells

تتركب كل شبكة عصبونية من مجموعة فيها n عنصر حساب عصبونـــي بسيط، $C=\{c_i\}, i=1,2,...,n$، وهناك ثلاثة أنواع من الخلايا الوظيفية: خلايا دخل، وخلايا خرج، وخلايا داخلية أو مخفية.

تتصل خلايا الدخل على نحو ما بمنبه خارجي (خلايا شمسية مثلاً) ينتج إشارات دخل متزامنة أو غير متزامنة. قد تكون إشارات الدخل قيماً حقيقية مستمرة أو مقطعة، سنرمز لها بالشعاع **x** ذي البعد **n**.

تُنتج خلايا الخرج إشارة خرج الشبكة وهي عبارة عن شعاع **y** ذي البعد **m**، وقد تكون قيماً حقيقية مستمرة أو مقطعة، وهي بدورها مرتبطة بمؤثرات خارجية؛ كأجهزة العرض أو أدوات الخرج الأخرى. عندما تتسلّم هذه الخلايا إشارات الدخل، تحسب قيماً معينة وتمررها

بعدئذ إلى خلايا أخرى تجري عليها معالجة تالية.

بوجه عام، تحسب هذه الخلايا قيماً متعلقة بمفاهيم كيفية أو التطبيق بين هذه المفاهيم. لإعطاء خواص الخلايا، من الضروري تعيين الطريقة التـــي تضم فيها إشارات الدخل إلى دخل الخلية، وكيف تحول الخلية دخلها *net*، والعلاقات التوقيتية بين المداخل والمخارج إذا كانت وثيقة الصلة بعمل الشبكة. مثلاً، قد تكون المداخل مركبة خطياً عموماً، كما هو مبين في الفصل الأول :

$$net = \mathbf{x}^T \mathbf{w} = \sum_i x_i w_i \qquad (1.2)$$

أو أحياناً كجداءات متصالبة مُثَقّلة لإشارات الدخل :

$$net = \sum w_{ij} \prod x_{i1} x_{i2}, \ldots, x_{ik} \qquad (2.2)$$

وهذا النـــوع من التركيب ينفذ في وحدات تســـمى وحدات مجموع الجداءات (Sigma-Pi Units).

بدلاً من التركيب بالمجموع المثقل أو بمجموع الجداءات المتصالبة المثقلة، قد تكون بعض المداخل عبارة عن قياس للمسافة d بين شعاع الدخل **x** وأوزان الخلايا **w** (d[**x** ,**w**]). في مثل هذه الحالات، يمكن أن توافق أوزان الخلية أشكالاً أولية (Prototypical) أو مراكز تجمعات أشعة أشكال الدخل. يمكن أن تتغير أيضاً المسافة المترية المستعملة في بعض الشبكات (مسافة هامنغ (Hamming)، المسافة الإقليدية (Euclidean)، نظيم الشعاع)، وسنشرح ذلك لاحقاً.

بعد أن عُرف الدخل *net* للخلية، يجب أن يتعين التحويل الوظيفي للخلية المنجز على الدخل؛ أي تابع تفعيل الخلية. وهذا يمكن أن يأخذ واحداً من الأشكال العديدة للتوابع المتزايدة المستمرة، كتلك الموضحة في الفصل الأول (الشكل 3.1) أو أشكالاً أخرى، سنشرح جميع هذه التوابع لاحقاً.

يمكن أن تكون قيم التفعيل حقيقية موجبة أو سالبة، وقد تكون ثنائية أو ثنائية القطبية، أو ذات أشكال أخرى.

أخيراً، من الضروري تعيين القيود الزمنية أو الشروط المفروضة على الحسابات ضمن الخلايا المختلفة وبينها.

2.2.2 نموذج التوصيل Connectivity pattern

توفر الوصلات الداخلية بين عناصر المعالجة واتجاهات تدفق الإشارة ضمن الشبكة المعلومات الأساسية لبنية النظام الأساسية. ويكون لخطوط الدخل المتصلة مع كل خلية C_i عادة وزنٌ مرافق W_{ij} وهو عبارة عن قياس للتأثير الذي تكتسبه الخلية C_i من خرج الخلية C_j أو من إشارات الدخل. سيكون للأوزان الموجبة تأثير تهييجي، على حين سيكون للأوزان السالبة تأثير تخميدي. وتوافق قيم الأوزان الصفرية عدم الربط (عدم وجود وصلة بين العقد).

تعيِّن الأوزان على الوصلات بين طبقتين أو مجموعتين من الخلايا بواسطة مصفوفة الوزن **W**. تعيِّن هذه المصفوفة كامل التوصيلات الداخلية للشبكة واتجاه انتشار الإشارة ضمنها. مثلاً، يُستعمل عنصر المصفوفة W_{ij} اصطلاحاً للدلالة على الوزن المتصل مع خرج الخلية j ودخل الخلية i (بعض المؤلفين يستعمل W_{ij} للدلالة على العكس). بهذا الاصطلاح نستطيع التعبير عن جهة انتشار الإشارات الأمامي أو الخلفي أو الجانبـــي أو التغذية العكسية - الذاتية. من أجل بعض بنـــى الشبكات العصبونية الصنعية العامة، يصبح من الأهمية بمكان تعيين الزمن أو/و الأمر الذي تنتشر فيه الإشارات بين الخلايا.

توافق قيم الأوزان على كل خطوط الوصلات الداخلية وسطاء الليف العصبي، وعليها تقوم مقدرة تخزين الشبكة، وبكلمات أخرى، المعرفة الموزعة في الشبكة. بوجه عام، يمكن أن تعدل الأوزان، وهي عبارة عن أعداد حقيقية، خلال عملية تعليم الشبكة أو عمليات أخرى.

3.2 حالة النظام System state

تحدد حالة النظام في أية لحظة مباشرة، بعد تعيين خواص الخلية نهائياً؛ يشمل ذلك قواعد تركيب الدخل، ونماذج الوصل الداخلي بين الخلايا، وتوابع التفعيل، وأمر انتشار الإشارة.

إن حالة النظام في لحظة زمنية ما، ولتكن t، هي قيم التفعيل (قيم الخرج) عبر مجموعة كل الخلايا في اللحظة t، وهذه القيم معينة بواسطة شعاع تابع التفعيل ذي البعد n، الذي

يحتوي على معلومات عما تمثله الشبكة عند لحظات مختلفة t.

4.2 عملية التعليم Learning process

التعليم المتكيف هو ميزة أساسية للأنظمة التـــي يجب أن تُوظَّف في أوساط متغيرة أو يجب أن تكون المكافئ للتغيرات في بنية الدخل (مثلاً، أشكال إشارة الكلام أو الكتابة اليدوية لأشخاص مختلفين).

وقد طُوِّرت إجراءات التعليم التكراري لبنـــى مختلفة. وهذه الإجراءات تتطلب قدرات حسابية هائلة في بعض أصناف الشبكات. وهكذا يصبح من الأهمية بمكان تطوير خوارزميات التعليم الفعالة، وخاصة من أجل الشبكات متعددة الطبقات وذات العدد الهائل من الوصلات الداخلية. هذا يفسر سبب اهتمام الباحثين بمسألة فعالية التعليم عبر السنين القليلة الماضية.

يـــنجز التعليم في الشبكات العصبونية الصنعية وفق إحدى الطرق التالية: بكيفية إنشاء الوصـــلات الداخلية بين العقد، أو بتعديل قيم الأوزان على الخطوط الرابطة للعقد (الموافق للـــتهييج أو الـــتخميد الليفي في الشبكات البيولوجية)، أو بتعديل قيمة عتبة توابع التفعيل للعقد، أو بتنفيذ هذه العمليات الثلاث بعضها مع بعض.

إذا كان دخل الانحياز مشمولاً بكل عقدة في الشبكة، وكان عدد العقد الأصلية والوصلات الداخلية كافياً للمسألة المعالجة، فإنه من الممكن تنفيذ التعليم بواسطة تعديل الأوزان فقط، باعتبار أن وزن الانحياز يعمل كقيمة عتبة، كما أن استعمال أوزان بقيم حقيقية بما في ذلك الصفر يساعد على تحقيق عمليات التهييج والتخميد وشروط عدم الربط بين العقد.

في هذه الحالة تكون مسألة التعليم، عموماً، عبارة عن عملية إيجاد قيم مصفوفة الوزن **W** المحققة للمعادلات الشعاعية التالية :

$$\mathbf{y}^p = \mathbf{F}(\mathbf{x}^p, \mathbf{W}) \qquad (3.2)$$

مهما يكن الدخل $\mathbf{x}^p$، حيث $p = 1,2,....,P$ ، و **F** تابع شعاعي غير خطي عموماً.

يحدث التعليم بتغير الأوزان بواحد من الأشكال العديدة؛ فالأوزان يمكن أن تعدَّل كتابع

لشدات إشارة الدخل، أو كتابع لكلا شدات إشارة الدخل والخرج، أو كتابع للأوزان نفسها، أو كتابع للفرق بين نموذج الخرج المنشود (target) المثالي والمحسوب (الخطأ)، أو تنفيذ كل هذه العمليات السابقة بعضها مع بعض. أيضاً، يمكن أن ينجز تعديل الأوزان على أساس الطرائق الاحتمالية.

5.2 خواص الشبكات العصبونية الصنعية

قبل البحث في قدرات كل من الشبكات العصبونية الصنعية على حده، من المفيد أن نتعرّف مميزاتها العامة وحدود إمكاناتها، وقدرتها على التطبيق العام، وما تستطيع (أو لا تستطيع) أن تتعلمه، وخواصّ أخرى مثل الوثوقية وسرعة المعالجة وغيرها. سنبدأ بتعرّف إمكاناتها في التطبيق.

1.5.2 مقدرات التطبيق Mapping capabilities

يمكن أن ينظر إلى الشبكة العصبونية الصنعية كمخطط صندوقي يحول أشعة الدخل **x** في الفراغ ذي البعد n إلى شعاع خرج **y** في الفراغ ذي البعد m :

$$\mathbf{F} : \mathbf{x} \rightarrow \mathbf{y}$$

بوجه عام، يكون التطبيق **F** إما مرافقاً ذاتـــياً (Autoassociative)؛ أي إعطاء الشكل الأصلي على الخرج من شكل الدخل الضجيجي أو الجزئي، أو مرافقاً مغايراً (Heteroassociative)؛ أي إعطاء نموذج خرج مختلف من نموذج الدخل.

2.5.2 التعليم والتعميم Learning and Generalization

إن التعميم هو عملية وصف للكل انطلاقاً من جزء أو بعض الأجزاء، أو تعليل حالة عامة من حالة خاصة، أو تعريف نوع من الأشياء انطلاقاً من معرفة عينة أو عدة عينات.

يعتبر التعميم جزءاً أساسياً من التعليم لأنه يسمح لنا بتذكر الحقائق الخاصة التـــي تطبق على كل الصفوف بدلاً من تذكر بعض الحقائق الخاصة التـــي تطبق فقط على أعضاء منفردة مـــن الصف. إذاً يخـــدم التعميم كنمط فعـــال للتذكر والتخزيـــن (Memorization/Storage). وبدون إمكانية التعميم، سنكون مضطرين لتذكر واستدعاء

عدد غير محدود من الحوادث الخاصة؛ كالحقائق والعلاقات وتفاصيل أخرى متعلقة بتجاربنا، وهذه مهمة غير ممكنة.

باختصار، التعميم هو معالجة أساسية للسلوك الذكي. تعمّم الشبكات العصبونية الصنعية عندما تحسب أو تستدعي أشكالاً كاملةً ابتداءً من أشكال دخل ضجيجية أو جزئية، أو عندما تنظم أو تصنف أشياء لم تتدرب عليها من قبل، أو عندما تتنبأ بنتائج من تصرفات ماضية.

إن مقدرة الشبكات العصبونية الصنعية على تصنيف الأشياء التـــي لم تتدرب عليها من قبل هو شكل من الاستيفاء الداخلي (Interpolation) بين الأشياء التـــي دربت عليها، والمقدرة على التنبؤ من تصرفات ماضية هو شكل من الاستيفاء الخارجي (Extrapolation). وكلا هذين النوعين من التطبيق هو شكل من التعميم.

في أنظمة حساب الذكاء الصنعي التقليدية (وتسمى أيضاً نظم الحساب الرمزية)، ينفذ التعميم من خلال تصنيف الأشياء؛ أي الصفوف المرغوب بها. يجزئ هذا التصنيف المجموعة الكلية للأشياء إلى صفين: أشكال منتمية إلى الصف المنشود وأشكال منتمية إلى الصف المتمم.

في هذه الأنظمة التقليدية، تعرّف هذه الصفوف بالتنبؤات التـــي تصف عادة أمثلة التدريب، وينجز التعميم بإحدى الطرائق العديدة التالية: الاستعاضة عن المتحولات بثوابت في التنبؤات، أو إهمال الحدود المترافقة (Conjunctive terms) في الواصفات (Descriptors)، أو إضافة حدود غير مترافقة في الواصفات، أو تســـلق شـــجرة التعميم، أو أية طريقة أخرى (Patterson عام 1993[3]).

مثلاً، من المراقبات التالية نجد:

طير (عصفور دوري) و طيران (عصفور دوري)

طير (أبو الحناء) و طيران (أبو الحناء)

طير (صقر) وطيران (صقر)

طير (كناري) وطيران (كناري)

طير (بلبل) وطيران (بلبل)

يستطيع المرء الآن أن يرتئي مفهوماً ضمنـــياً معمماً هو: **كل الطيور تطير**. مهما يكن x، إذا كانت x طيراً، فإن x يستطيع الطيران، كما يلي:

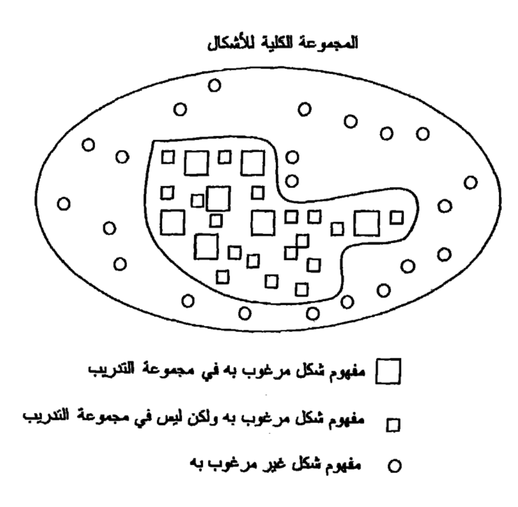
$$\forall x \quad \text{bird}(x) \Rightarrow \text{fly}(x) \qquad (4.2)$$

لاحظ في هذه الحالة أن التعميم أنجز بإبدال المتحول x بالثوابت: عصفور دوري، أبو الحناء، ...الخ. وهكذا حتـــى نحصل على القاعدة المستنتجة: **كل الطيور تطير**.

بالطبع إن التعليم المستشف ليس شاملاً أو راسخاً في الإدراك المنطقي، لذا قد يكون هذا المفهوم خاطئاً، فمثلاً من طير (نعامة) والمفهوم الضمنـــي المذكور آنفاً نستنتج طيران (نعامة). ومع أن هذا الاستنتاج مضحكٌ وليس شكلاً صحيحاً منطقياً للتعليل، إلا أنه نافع في كلٍّ من مهام التعليل والتعميم، كما سنرى لاحقاً.

ينجز التعميم في الشبكات العصبونية الصنعية أيضاً بواسطة تشكيل الصفوف. وفي هذه الحالة، تعرَّف الصفوف بتابع التطبيق $\mathbf{F}$ كحدود في فراغ الأوزان. بعض الحدود تحيط بقيم فراغ هيئة (مَعْلَم) الشيء التـــي تحقق الصفوف المنشودة والصفوف المتممة. تشكل الحدود من خلال عملية تعديل الوزن الليفي في عمليات التعليم. نقول إن التعليم أو التعميم دقيق عندما تضم حدود الصف النهائي فقط الأمثلة المنشودة وتستبعد كل الأمثلة غير المرغوب بها. نعنـــي بالأمثلة المنشودة كل الأشكال المنتمية، انتماءً صحيحاً، إلى صفوف المفهوم المعطى، بما في ذلك، الأمثلة المستعملة في مجموعة التدريب وغير المستعملة، كما هو موضح في (الشكل 1.2).

نستطيع تعريف التعميم بالطريقة التالية: لتكن لدينا أشعة دخل معطاة $\mathbf{x}^p$، حيث $p=1,2,...,P$ ، ذات توزيع احتمالي $\rho(\mathbf{x})$، وليكن $\mathbf{y}^p = \mathbf{f}'(\mathbf{x}^p)$ هو خرج الشبكة المحسوب لمجموعة معطاة من أوزان الشبكة $\mathbf{w}$. نريد من التابع f'، بتطبيق صحيح مرغوب به، أن يكون قريباً من f أقرب ما أمكن من أجل كل قيم $\mathbf{x}$ (ليس بالضبط عند نقاط مجموعة التدريب $\{(\mathbf{x}^p, \mathbf{t}^p) \quad | p=1,2,\cdots,P\}$)، حيث $\mathbf{t}^p$ خرج الشبكة المرغوب به، أو المنشود، الموافق للدخل $\mathbf{x}^p$ في مجموعة التدريب.



الشكل 1.2: تعليم مفهوم معمم

3.5.2 الإنجاز الموثوق Robust performance

يمكن اعتبار الشبكات العصبونية الصنعية كأنظمة حساب موثوقة؛ فهي تستمر بالإنجاز جيداً عندما يكون جزء من الشبكة ضعيفاً أو يمثل بمعطيات ضجيج. وهذا ممكن لأن المعرفة (Knowledge) المخزنة في الشبكات العصبونية الصنعية موزعة عبر الكثير من العصبونات والوصلات الداخلية وليس عبر وحدة واحدة أو بضع وحدات. أي أن المفاهيم أو التطبيقات المخزنة في الشبكة لها درجة بناء زائدة نوعاً ما عن الحاجة من خلال هذا التوزيع للمعرفة.

هذا الموضوع في الشبكات العصبونية الصنعية يدعى أحياناً مجال التسامح بالخطأ (fault tolerance). وهو مغاير تماماً للحواسيب التقليدية، إذ إن ضياع ترانزستور واحد، أو عطل أي أداة أخرى في حاسوب تسلسلي يعمل بمبدأ فون نيومان، يمكن أن يؤدي إلى إخفاق النظام بالكامل.

إن بعض الأنظمة غير متسامحة كثيراً بالخطأ. ويستطيع المرء توقع إبداء الدماغ البشري خواص مشابهة للشبكات العصبونية الصنعية. وبالفعل فهذه هي الحالة؛ حيث يمكن أن يتضرر جزء من الدماغ (أو يُزال) دون حصول تأثيرات خطيرة على أداء الشخص. إذاً المعالجة الفعالة بمعطيات ضجيجية هي معيار عند الإنسان، وليس حالة استثنائية.

4.5.2 المعالجة المتوازية Parallel processing

إن الدماغ والشبكات العصبونية الصنعية قادرة على معالجة المعطيات بسرعة عالية حقاً. لنتصور الآن كمية الحسابات اللازمة لمعالجة صورة مرئية واحدة. فإذا أخذنا فقط 1000 × 1000 مستقبل لمعالجة صورة مرئية، مع العلم أن هذا الرقم صغير إذا ما قورن بشبكية العين، حيث أكثر من مليون رقم (ثلاثة ملايين للصورة المرئية الملونة) يجب أن تفحص لكي تكون الأشياء في الصورة مميزة، ويجب أن تنجز عدة ملايين عملية حسابية.

هذه المهمة تتطلب عدة ثوان في الحاسوب التقليدي، حتى عند سرعات من رتبة 10^{-9} ثانية للحاسوب الحديث، أي إن التمييز يمكن أن يكون هامشياً (marginal). وبالمقابل، تحسب مثل هذه المهام باستخدام الأجهزة البصرية البيولوجية بزمن من رتبة الميللي ثانية، مع العلم أن الإشارات تنتشر بمعدل ميللي ثانية في الجهاز البيولوجي. ومردّ هذا التباين في الإنجاز، يعود إلى التنفيذ اللحظي للحسابات التفرعية الهائلة ضمن الشبكة البيولوجية.

على الرغم من أن الشبكات العصبونية الصنعية تشترك في خواص عديدة مع الشبكات البيولوجية، إلا أنها مازالت عاجزة عن بعض الخواص كالنسيان مثلاً (Zak عام 1991[4]).

6.2 لمحة عن طرائق تعليم الشبكات العصبونية الصنعية

يمكن أن تصنف طرائق تعليم الشبكات العصبونية الصنعية في ثلاثة أنواع أساسية: مع معلم (Supervised)، وتقوية (Reinforcement)، وبدون معلم (Unsupervised).

في التعليم مع معلم (Teacher)، من المفترض وجود أستاذ خلال عملية التعليم، ويشمل كل نموذج يُستعمل لتدريب الشبكة شكل الدخل مع شكل الخرج الموافق المنشود أو المرغوب به(الجواب الصحيح). خلال عملية التعليم، يمكن أن تجري المقارنة مع الخرج المحسوب بواسطة الشبكة والخرج المنشود وذلك لتعيين الخطأ. يمكن أن يستعمل الخطأ بعدئذ

تغيير وسطاء الشبكة لتعطي تحسيناً في الإنجاز.

عادة تعطى مصفوفة الأوزان الواصلة بين الطبقات قيماً بدائية مساوية للصفر، أو أعداداً قيم حقيقية عشوائية صغيرة جداً (قيم كيفية). عندما تغذى أشعة شكل تدريب الدخل x^p، حيث $p = 1,2,\ldots,P$ ، من الشبكة في لحظة زمنية ما، ستقوم الشبكة بحساب شعاع شكل الخرج الموافق y^p مباشرة. يقارن شكل الخرج المحسوب مع شكل الخرج المنشود t^p ويعطى الخطأ بالعلاقة التالية :

$$e^p = y^p - t^p \qquad (5.2)$$

يستعمل الخطأ الناتج بعدئذ من خلال شكل ما من الحسابات والتغذية العكسية ليعدل الأوزان منفردة، وذلك لتقليل الخطأ من أجل كل زوج تدريب. بعد التعديل المتكرر للأوزان لكل أشكال التدريب، يمكن أن تتقارب قيم الأوزان إلى مجموعة من القيم الضرورية لإنجاز استدعاءات الشكل المطلوب.

ينجز التعليم عندما تكون الأخطاء لكل نماذج التدريب $(p = 1, 2,\ldots,P)$ صغيرة حتـــى المستوى المقبول لكل الأشكال الجديدة التـــي ليست في زمرة التدريب.

وفي تعليم التقوية أو التعزيز، من المفترض وجود الأستاذ أيضاً، لكن الجواب الصحيح لا يكون في الشبكة. وبدلاً من ذلك، تمثل الشبكة فقط بمؤشر فيما إذا كان جواب الخرج صحيحاً أو خاطئاً. بعدئذ يجب أن تستعمل الشبكة هذه المعلومات لتحسين الإنجاز.

عادة يكون هناك مكافأة تعطى بواسطة تقوية قيم الأوزان على الوحدات التـــي أعطت جواباً صحيحاً، وعقوبة بإنقاص قيم الأوزان تفرض على الوحدات التـــي أعطت جواباً خاطئاً.

وفي التعليم بدون معلم، ليس لدى الشبكة تغذية عكسية على الجواب الصحيح أو المرغوب به، وليس هناك أستاذ مندوب يمثل الأشكال المنشودة، لذا يجب على النظام أن يتعلم بواسطة اكتشاف وملائمة الخصائص أو الحقائق البنيوية في أشكال الدخل، هذا يعنـــي الملائمة مع قواعد إحصائية أو تجمعات الأشكال من عينات تدريب الدخل.

يمكن أن ينجز مثل هذا التعليم بواسطة تقوية حساسيات العقد المختارة (الأوزان) لتتلائم مع أشكال التدريب الأولية المركزية والممثلة لمجموعة الأشكال المتشابهة أو التجمعات.

وُجدتْ أمثلة عن خوارزميات التعليم هذه في شبكة هامنغ (Lipmann عام 1987[5])، ونموذج Linsker عام 1988[6]، ونموذج Kohonen عام 1982[7]، ونماذج cognitron و neocognitron لليابانـــي Fukushima عام 1982[8]، ونماذج نظرية الطنين المتكيف ART (Adaptive Resonance Theory) للباحث Grossberg عام 1988[9].

ينجز التعليم في هذه الشبكات بتقوية الأوزان على خطوط الوحدات ذات الاستجابة الضعيفة، أما الوحدات ذات الاستجابة القوية فيمكن أن تخمد مخارج وحدات أخرى موجودة على نفس الطبقة، يحدث هذا في شبكات: (neocognitron, Kohonen, Hamming)، أو يمكن أن تغذى التخميدات عكسياً بين الطبقات المتتالية، (ويحدث هذا في شبكات نظرية الطنين المتكيف (ART)).

هناك صنف هام من طرائق التعليم بدون معلـــم، يعرف بالتعليم التنافســـي (competitive learning).

ومع أن تعليم التقوية دُرس من قبل الباحثين على نطاق واسع، فإنه ليس من أشكال التعليم الشائعة كثيراً، لذا سنحصر اهتمامنا بالتعليم بمعلم وبدون معلم فقط.

1.6.2 التعليم الترابطي أو قاعدة Hebb للتعليم

Hebbian or Correlative learning

تعليم Hebb هو شكل من تعديل الأوزان الترابطي (مرتبطة مع قوة العصبون الليفية السابقة واللاحقة). اقترح هذه النظرية الأساسية Hebb عام 1949[1]، واستُخدمت للتعليم في شبكات الذاكرة المترافقة وحيدة الطبقة وفي الشبكات التـــي تتعلم بدون معلم.

ينجز التعليم في هذه الحالة بطريقة مباشرة، حيث تستعمل أشكال الدخل $\mathbf{x}^p$ وأشكال الخرج المنشـــود الموافقة لها $\mathbf{t}^p$ لحســـاب مصفوفة الوزن $\mathbf{W}$ كمجموع $\mathbf{w}^p$ مصفوفة شـــكل $(p=1,2,....,P)$، حيث تحسب كل مصفوفة $\mathbf{w}^p$ كجداء خارجي أو مصفوفة ترابط:

$$\mathbf{W}^p = \mathbf{x}^p(\mathbf{t}^p)^T \quad , \quad \mathbf{W} = \sum_{p=1}^{P} \mathbf{W}^p \qquad (6.2)$$

ففي حالة المداخل *net* لشبكة الترافق الذاتـــي، مثل شبكة Hopfield المصممة عام

198[10]، وشبكة حالة – دماغ – في- صندوق (BSB) لـ Anderson المقترحة عام 198[11]، سيكون لدينا $t^p = x^p$.

وفي حالة مداخل *net* لشبكات ذواكر الترافق المغاير مثل الذاكرة المترافقة ثنائية الاتجاه لـ Kosko المقترحة عام 1987[12] وشبكات هامنغ (Hamming) لـ Lippman المقدمة عام 1987[5]، فإن الأشعة x^p و t^p ستكون مختلفة بعضها عن بعض عموماً.

لقد اقتُرحت طرائق متنوعة كثيرة تطبق قاعدة Hebb، يشمل ذلك تعديل الأوزان المبنـــي على تقليل الطاقة أو تابع الأنتروبـــي (سنشرح كل ذلك بالتفصيل لاحقاً).

2.6.2 التعليم الاحتمالي Stochastic learning

يجري في هذا النوع من التعليم تعديل الأوزان بطريقة احتمالية. وقد وُجدت الأمثلة عن هذا التعليم في محاكاة التلدين (Simulated annealing) كالذي طبق في آلات كوشي وبولتزمان (Cauchy وBoltzmann) حيث تعيَّن حالات كل الوحدات بواسطة توزيع احتمالي.

يعمل النظام خلال طور التعليم وفق نمطين: نمط الإلزام (Clamped mode)؛ حيث تكون عقد الدخل/ الخرج ملزَمة بقيم من الأزواج المرافقة للأشكال الثنائية، ونمط عدم الإلزام (Unclamped mode)، حيث لا توجد مداخل. في هذا النظام يُسمح للشبكة بالعمل وفق كلا النمطين حتـــى تصل إلى التوازن الحراري؛ وذلك في اللحظة التـــي تكون عندها قد عدلت الأوزان w_{ij} (ذات القيم الحقيقية على الوصلات بين العقد C_i و C_j) على أساس الفرق بين حالتين احتماليتين $P_{ij}(U)$ و$P_{ij}(C)$. حيث $P_{ij}(U)$ هو الاحتمال الذي يكون فيه كلا العنصرين C_i وC_j في النمط الملزم، و$P_{ij}(C)$ هو الاحتمال الذي تكون فيه كلـــتا الوحدتين في النمط غير الملزم.

من الممكن تعريف النظام عندما تكون الشبكة متناظرة بالعلاقة التالية:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}s_i s_j \qquad (7.2)$$

حيث s_i هو مقدار ثنائي (0,1) أو ثنائي القطبية (+1,−1) للوحدة c_i. تصل الشبكة إلى التوازن عندما يصل تابع الطاقة إلى قيمته الدنيا.

3.6.2 تعليم تدرج الهبوط Gradient descent

هناك عدة طرائق للتعليم مبنية على إنقاص الخطأ أو تابع الكلفة E، وذلك من خلال استعمال طرائق تدرج الهبوط. تتطلب هذه الطرائق أن تكون توابع التفعيل تفاضلية عندما يكون تغير الأوزان مبنياً على تدرج الخطأ المعرف بعبارات الأوزان وتوابع التفعيل.

يعطى شكل قاعدة التحديث كحل للمعادلة التالية:

$$\Delta w_{ij} = \eta \frac{\partial E}{\partial w_{ij}} \qquad (8.2)$$

حيث η وسيط معدل التعليم، و w_{ij} الوزن على الوصلة بين الوحدة i والوحدة j.

يشمل هذا النوع من التعليم قاعدة دلتا لـ Widrow-Hoff وخوارزمية الانتشار الخلفي الشائعة (BP) (ستناقش بالتفصيل جميع هذه القواعد لاحقاً).

إن إجراء التعليم بقاعدة دلتا يكون بالمفهوم الأمثلي في الشبكات العصبونية الصنعية وحيدة الطبقة. وستوجد هذه القاعدة مصفوفة الأوزان التـي تعطي استدعاءً تاماً عندما تكون أشكال الدخل مستقلة خطياً، أو بوجه عام، للأشكال التـي تبدي ارتباطات عالية بمفهوم المربعات الصغرى.

على أية حال، أخفق الإجراء في إعطاء حل عند استخدامه في الشبكات متعددة الطبقات، إذ إنه لا يستطيع تعيين كيفية تعديل أوزان الوحدات في الطبقات الداخلية خلال عملية التعليم. وهذا يعرف بمشكلة تعيين الاعتماد (Credit-assignment). إذ ليس هناك وسيلة واضحة لتعيين الاعتماد أو المسؤولية (blame) لأوزان وحدة الطبقة الداخلية من أجل تخفيض الأخطاء في وحدات الخرج.

لقد حُلّت مشكلة تعيين الاعتماد باستعمال شبكة بطبقات متعددة تَستعمل توابع تفعيل مختلفة، وتقوم بالتعديل التكراري لأوزان الطبقة الداخلية باستعمال الأخطاء المحسوبة من طبقة الخرج والانتشار الخلفي لتعديل الأخطاء، طبقة بعد طبقة، حتـى تصل إلى أول طبقة داخلية

ن طرف الدخل.

ليس التقارب مؤكداً في خوارزمية الانتشار الخلفي كما هو مؤكد في قاعدة دلتا، حيث ن العملية يمكن أن تجد لها نهاية صغرى محلية (local minimum) حيثما توقفت، وسنشرح ذلك بالتفصيل لاحقاً.

7.2 تصنيف الشبكات العصبونية الصنعية

في هذا المقطع الأخير من هذا الفصل سنقدم ثلاثة أصناف مختلفة للشبكات العصبونية الصنعية. سنبدأ باستراتيجية التعليم التـــي تشمل أصناف التعليم الموسعة: مع معلم، والتقوية وبدون معلم، كما هو موضح في (الأشكال 2.2 و3.2).

ففي كل فئة من هذه الفئات، تقع الشبكة العصبونية الصنعية ضمن واحدة من أربع فئات جزئية، ذُكرت جميعها في (الشكل 4.2).

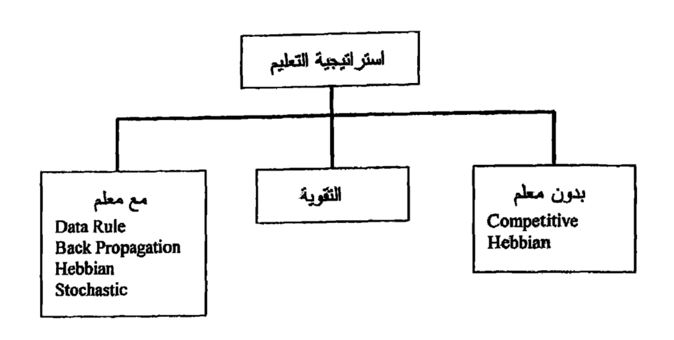
ستعرض أيضاً للتصنيف وفقاً لنوع بنية الشبكة، وهذا موضح في (الشكل 5.2).

وُضِع التصنيف الأخير للشبكات على أساس نوع التطبيق (المسألة المعالجة)، حيث صنفنا الشبكات في ست مناطق تطبيقية، كما هو موضح في (الشكل 6.2).

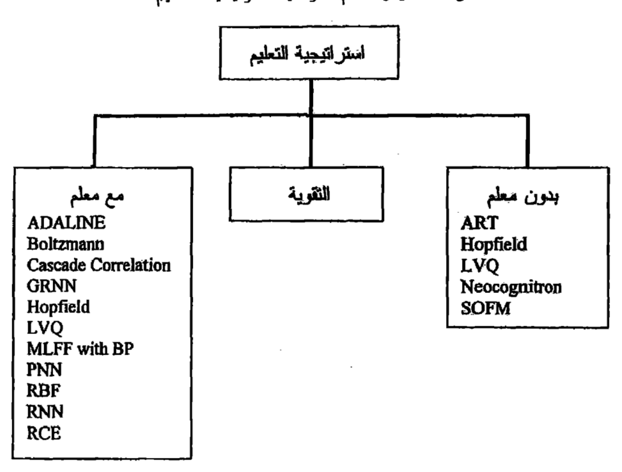
وبغية سهولة التمثيل لجميع الأصناف بالشكل المفصل، آثرنا ذكر الرموز العلمية المختزلة لأسماء كل الشبكات ضمن الأشكال السابقة، وسنقدم شرحاً لهذه الرموز في الجدول التالي:

ADALINE: ADAptive LInear Neural Element	عنصر عصبونـــي خطي متكيف
ART: Adaptive Resonant Theory	نظرية الطنين المتكيف
AM: Associative Memories	ذواكر مترافقة
BAM: Bidirectional Associative Memory	ذاكرة ترافق ثنائي الاتجاه
BM: Boltzmann Machine	آلة بولتزمان
BSB: Brain-State-in-a-Box	حالة دماغ في صندوق
CCN: Cascade Correlation Networks	شبكات الارتباط المتتابع
CM: Cauchy Machine	آلة كوشي

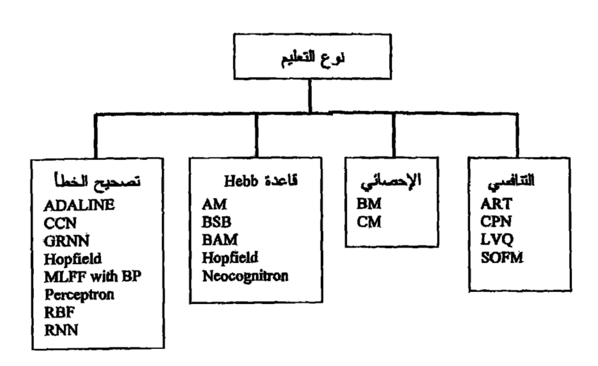
CPN: Counter Propagation Network	شبكة الانتشار المتعاكس
GRNN: Generalized regression Neural Network	شبكة عصبونية تراجعية معممة
Hamming Network	شبكة هامنغ
Hopfield Network	شبكة هوبفيلد
LVQ: Learning Vector Quantization	التعليم بالتكميم الشعاعي
MADALINE: Many ADALINE	عناصر عصبونية خطية متكيفة
MLFF: MultiLayer FeedForward Backpropagation Network	شبكة الانتشار الخلفي متعددة الطبقات بتغذية أمامية
Neocognitron	
NLN: NeuroLogic Networks	شبكات المنطق العصبوني
Perceptron	المقسر
PNN: Probabilistic Neural Network	شبكة عصبونية احتمالية
RBF: Radial Basis Function	تابع الأساس الشعاعي
RNN: Recurrent Neural Networks	شبكات عصبونية تكرارية
RCE: Reduced Coulomb Energy	طاقة كولومب المخفضة
SOFM: Self-Organizing Feature Map	خريطة الملامح ذاتية التنظيم



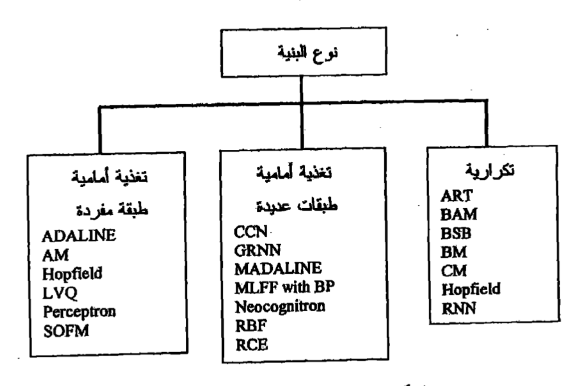
الشكل 2.2: طرائق التعليم النموذجية لاستراتيجيات التعليم المختلفة



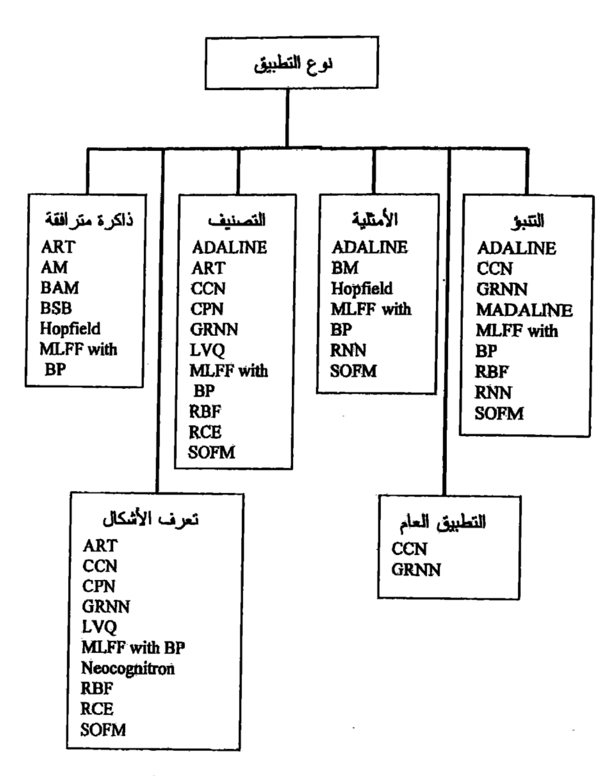
الشكل 3.2: فئات أنواع الشبكات حسب طريقة التعليم



الشكل 4.2 فئات أنواع الشبكات حسب طرق التعليم



الشكل 5.2: فئات الشبكات حسب نوع البنية



الشكل 6.2: فئات أنواع الشبكات حسب نوع التطبيق

مراجعة عامة لبعض المفاهيم الرياضية والإحصائية

سنشرح في هذا الفصل بعض المواضيع المختارة في الرياضيات والإحصاء الضرورية لفهم الشبكات العصبونية الصنعية. سنهتم خصوصاً بمراجعة الجبر الخطي (أشعة، مصفوفات)، والحساب التفاضلي والتكاملي، والاحتمالات والإحصاء، ونظرية المعلومات والمنطق العائم (الغامض/المبهم) (fuzzy logic).

سنتعرض أيضاً لنظرية الأنظمة غير الخطية والفوضوية (chaos) اللازمة لدراسة الشبكات العصبونية الصنعية التكرارية. يعتبر هذا العرض الموجز للمواضيع السابقة ضرورياً على امتداد هذا الكتاب.

1.3 تمهيد

لا يمكن أن تُقدِّر المقدِّرات والقيود والسلوكيات العامة لأنظمة الشبكة العصبونية حق التقدير بدون فهم الوظيفة الأساسية والديناميكية لبعض الأنظمة. وهذا يتطلب بالطبع فهماً عاماً لبعض النظريات الرياضية الأساسية، لذا سنقوم في هذا الفصل بمراجعة مختصرة لمواضيع منتقاة في الرياضيات والإحصاء تفيد أولئك الذين يريدون إنعاش معرفتهم في هذه المسائل من خلال تقديم التعاريف المبسطة والصيغ الأساسية لبعض المفاهيم الرياضية الخاصة.

القصد من معالجتنا لهذه المواضيع هو **الاختصار والشرح الكافي واللازم** لفهم أفضل لمفاهيم وأفكار الشبكات العصبونية الصنعية التـــي يتضمنها هذا الكتاب.

في البداية، سنقبل جميع المتحولات في هذا الفصل كأسماء وتعابير رياضية مجردة دون إقحامها في موضوع الشبكات العصبونية الصنعية حتـــى يتسنـــى لنا تعلم وفهم الأداة الرياضية أولاً، ومن ثم سيكون لكل مقام في الشبكات العصبونية مقال.

2.3 مراجعة للجبرالخطي والفراغ الشعاعي

Review of linear algebra and vector space

1.2.3 الأشعة Vectors

يُعرَّف الشعاع **x** بأنه عمود مؤلف من *n* عدداً كما يلي :

$$\mathbf{x}=\begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_n \end{pmatrix} \in \mathbf{R}^n \qquad (1.3)$$

حيث يمكن أن ينظر له كتعريف نقطة في الفراغ ذي البعد *n*.

يرمز لمنقول الشعاع **x** بـ $\mathbf{x}^T$، ويحصل عليه بتدوير شعاع العمود **x** ليشكل شعاع سطر من الأعداد كما يلي:

$$\mathbf{x}^T=(x_1,x_2,...,x_n) \qquad (2.3)$$

يعرف الجداء السلمي (scalar product) لشعاعَي سطر **x** و **y** بكمية سلمية كالآتـي:

$$\mathbf{x}^T\cdot\mathbf{y}=x_1y_1+x_2y_2+...+x_ny_n=\sum_i x_iy_i \qquad (3.3)$$

لتكن **x** و **y** و **z** أشعة، وليكن **0** شعاعاً صفرياً (كل قيمه أصفار) وليكن $c\in\mathbf{R}$ عدداً ما.

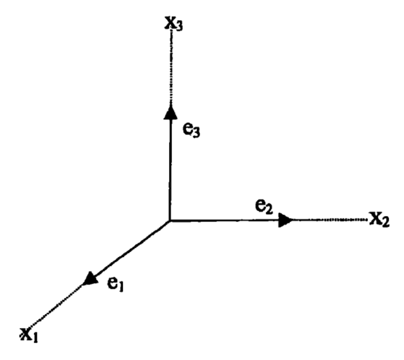
عند ذلك يمكننا كتابة خواص العمليات الشعاعية التالية:

$$\begin{aligned} &\mathbf{x}.\,\mathbf{y}=\mathbf{y}.\,\mathbf{x} \\ &\mathbf{x}.\,(\mathbf{y}+\mathbf{z})=\mathbf{x}.\,\mathbf{y}+\mathbf{x}.\,\mathbf{z}=(\mathbf{y}+\mathbf{z}).\mathbf{x} \\ &(c\,\mathbf{x}).\,\mathbf{y}=c\,(\mathbf{x}.\,\mathbf{y}) \qquad (4.3) \\ &\mathbf{x}.\,(c\,\mathbf{y})=c(\mathbf{x}\,.\mathbf{y}) \end{aligned}$$

إذا كان $\mathbf{x}=\mathbf{0}$ فإن $\mathbf{x}.\,\mathbf{x}=0$ وإلا $\mathbf{x}.\,\mathbf{x}>0$

نقول عن شعاعَي سطر غير صفريين **x** و **y** إنهما متعامدان إذا كانت قيمة جدائهما السلمي تساوي الصفر، أي $\mathbf{x}^T.\,\mathbf{y}=0$. مثلاً، الأشعة الواحدية e_1, e_2, e_3 المعطاة بالقيم

التالية: (1, 0, 0)، (0, 0, 1)، (0, 1, 0) على الترتيب هي أشعة متعامدة مثنى، كما هو موضح في (الشكل 1.3).



الشكل 1.3: الأشعة الواحدية المتعامدة في فراغ ثلاثي البعد

نظيم (أو طويلة) الشعاع **x** هو كمية سلّمية تعطى بالعلاقة :

$$\|\mathbf{x}\| = (\mathbf{x}^T \cdot \mathbf{x})^{1/2} = (\sum_{i=1}^{n} x_i^2)^{1/2} \qquad (5.3)$$

ويمكن استنتاج الخواص التالية بسهولة:

ـــ المسافة بين الشعاعين **x** و**y** تعطى بالعلاقة:

$$\|\mathbf{x} - \mathbf{y}\| = \left[(\mathbf{x} - \mathbf{y}) \cdot (\mathbf{x} - \mathbf{y})\right]^{1/2} \qquad (6.3)$$

ـــ إذا كان $\|\mathbf{x}\| = 1$ فإن **x** هو شعاع واحدي، ومن ثم إذا كان $\|\mathbf{x}\| = a$ فإن $(1/a)\mathbf{x}$ هو شعاع واحدي.

ـــ $\|c\,\mathbf{x}\| = |c|\,\|\mathbf{x}\|$، حيث | | القيمة المطلقة

ـــ إذا كان $c\,\mathbf{x} = \mathbf{y}$، حيث $c > 0$، فإن **x** و**y** لهما نفس الاتجاه. وهكذا $\mathbf{x}/\|\mathbf{x}\|$ هو شعاع واحدي في نفس الاتجاه كما هو **x** $(\mathbf{x} \neq 0)$.

مسقط الشعاع **x** على طول شعاع آخر **y** هو الشعاع **c y** المعطى بالعلاقة التالية:

$$c\mathbf{y} = \mathbf{y}\frac{(\mathbf{x}\cdot\mathbf{y})}{(\mathbf{y}\cdot\mathbf{y})}$$

و تعطى الزاوية θ بين الشعاعين **x** و**y** بالعلاقات:

$$\mathbf{x}\cdot\mathbf{y} = \|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta \quad \Rightarrow \quad \cos\theta = \frac{\mathbf{x}\cdot\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} \qquad (7.3)$$

هناك علاقات هامة بين أي شعاعين **x** و**y** تعطى بمتراجحة كوشـــي ـــ شوارتز (Cauchy-Schwartz) التـــي تنص على أن مربع الجداء السلمي لشعاعين هو أقل أو يساوي جداء مربع نظيمي الشعاعين:

$$(\mathbf{x}\cdot\mathbf{y})^2 \le \|\mathbf{x}\|^2\,\|\mathbf{y}\|^2 \qquad (8.3)$$

نقول عن مجموعة الأشعة $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ إنها مرتبطة خطياً، إذا وُجدت مجموعة من الأعداد $(c_1, c_2, \ldots, c_n)$ ليست كلها أصفاراً بحيث $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_n\mathbf{x}_n = 0$، وإذا لم يوجد مثل هذه الأعداد نقول عن مجموعة الأشعة هذه إنها مستقلة خطياً. بكلمات أخرى، تكون مجموعة الأشـــعة $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ مســـتقلة خطياً إذا وفقط إذا كان مجموع الجداءات $\sum_i c_i\mathbf{x}_i = 0$ [مهما كانت الثوابت $(c_1, c_2, \ldots, c_n)$] فإن $c_i = 0$ [مهما كانت القيم $i = 1,2,\ldots, n$].

في الفراغ $\mathbf{R}^n$، هناك على الأكثر n شعاع مستقل خطياً. وأكثر من ذلك، أي n شعاع مستقل خطياً في $\mathbf{R}^n$ يمكن أن يولّد الفراغ $\mathbf{R}^n$. وهذا يعنـــي أن أي شعاع في $\mathbf{R}^n$ يمكن أن يعبَّر عنه (أو يولَّد) كتركيب خطي لأشعة مستقلة خطياً. مثل هذه الأشعة يقال عنها إنها تشكل أساس الفراغ $\mathbf{R}^n$.

2.2.3 المصفوفات Matrices

تتألف المصفوفة من m سطراً و n عموداً. سنرمز للمصفوفات بحروف كبيرة، مثل **A** أو بواسطة $m \times n$ عنصر مصفوفة a_{ij}، حيث $i = 1,2,\ldots,m$، $j = 1,2,\ldots n$.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ . & . & . & . \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \qquad (9.3)$$

أو ببساطة بالتدوين المختصر $\mathbf{A} = (a_{ij})$.

المصفوفة الصفرية **O** هي مصفوفة جميع عناصرها أصفار.

المصفوفة الواحدية ويرمز لها **I** هي مصفوفة عناصرها على القطر الرئيسي ذات قيمة مساوية للواحد والبقية أصفار:

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ . & . & . & . \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad (10.3)$$

لاحظ أن الأعمدة (الأسطر) في المصفوفة الواحدية هي أشعة مستقلة خطياً.

يتحقق جمع (أو طرح) مصفوفتين فقط عندما تكونان من نفس الأبعاد. إذا كانت **A** و**B** مصفوفتين فإن مجموعهما (أو طرحهما) هو حاصل جمع (أو طرح) قيم العناصر ذات الأدلة المتماثلة $\mathbf{A}+\mathbf{B}=(a_{ij}+b_{ij})$.

الشعاع العمود هو مصفوفة خاصة ببعد m × 1 مؤلفة من عمود مفرد يحوي m عنصراً. والشعاع السطر هو مصفوفة خاصة ببعد 1 × n مؤلفة من سطر واحد يحوي n عنصراً.

يمكن تنفيذ بعض العمليات على المصفوفات مثل ضرب المصفوفة **A** بثابت c، كما يلي:

$$c\mathbf{A} = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ . & . & . & . \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix} \qquad (11.3)$$

إذا كان لدينا المصفوفة **A** ببعد m × n والمصفوفة **B** ببعد n × s، فإن مصفوفة الجداء **C = A × B** ستكون ببعد m × s والعنصر c_{ik} منها يعطى بالعلاقة:

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1} b_{1k} + a_{i2} b_{2k} + \dots + a_{in} b_{nk} \qquad (12.3)$$

وإذا كانت $\mathbf{A}_m, \dots, \mathbf{A}_2, \mathbf{A}_1$ الأشعة السطرية للمصفوفة **A** و$\mathbf{B}^s, \dots, \mathbf{B}^2, \mathbf{B}^1$ الأشعة العامودية للمصفوفة **B**، فإن العنصر *ik* من مصفوفة الجداء AB يساوي $\mathbf{A}_i\mathbf{B}^k$. لذا فإن مصفوفة الجداء الناتجة يمكن أن تكتب كما يلي:

$$\mathbf{C}=\mathbf{AB}=\begin{bmatrix} A_1B^1 & A_1B^2 & \dots & A_1B^s \\ A_2B^1 & A_2B^2 & \dots & A_2B^s \\ . & . & . & . \\ A_mB^1 & A_mB^2 & \dots & A_mB^s \end{bmatrix} \qquad (13.3)$$

وكذلك نحصل على منقول المصفوفة A، ويرمز له بـــ $\mathbf{A}^T$، بتبديل الأعمدة والأسطر بعضها مع بعض، بحيث يُدوَّر أول عمود ليكون محل السطر الأول، والعمود الثانـــي محل السطر الثانـــي وهكذا. فإذا كانت $\mathbf{A}=(a_{ij})$ مصفوفة ببعد $m \times n$ فإن $\mathbf{A}^T=(a_{ji})$ ستكون ببعد $n \times m$ وتكتب بالشكل التالي:

$$\mathbf{A}^T=\begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ . & . & . & . \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{bmatrix} \qquad (14.3)$$

إذا كانت المصفوفة مساوية لمنقولها، $\mathbf{A}=\mathbf{A}^T$، فإنها تسمى بالمصفوفة المتناظرة ويكون $a_{ij}=a_{ji}$.

يرمز لمقلوب أو عكس المصفوفة بـــ $\mathbf{A}^{-1}$، ويكون لدينا $\mathbf{A}^{-1}.\mathbf{A}=\mathbf{A}.\mathbf{A}^{-1}=\mathbf{I}$. من الواضح أن المصفوفة A يجب أن تكون مصفوفة مربعة $(n \times n)$ ليكون لها مصفوفة عكسية. بالطبع ليس كل المصفوفات لها عكس، لكن إذا كان هناك مصفوفة لها عكس، سيكون هذا العكس وحيداً، ويحقق: $(A^{-1})^T=(A^T)^{-1}$.

رتبة rank المصفوفة هي عدد أشعة الأعمدة المستقلة خطياً، فإذا كانت r رتبة المصفوفة، فإن المصفوفة سيكون لها r عمود مستقل خطياً وأيضاً r سطر مستقل خطياً.

يمكن أن ينظر للمصفوفة كتطبيق خطي $\mathbf{M}:\mathbf{R}^n \rightarrow \mathbf{R}^m$ من فراغ ببعد n إلى فراغ ببعد m. مثلاً، إذا كانت A مصفوفة ببعد $m \times n$، و $\mathbf{x}$ مصفوفة ببعد $n \times 1$ (شعاع عمود)، فإن الجداء $\mathbf{A}\,\mathbf{x}=\mathbf{y}$ سيكون مصفوفة ببعد $m \times 1$ (شعاع عمود)، هذا تطبيق من $\mathbf{R}^n$ إلى $\mathbf{R}^m$.

الجداء الخارجي (outer product) لشعاعين هو نوع خاص من التطبيق. فإذا كان لدينا مصفوفة $\mathbf{x}$ ببعد $n \times 1$ ومصفوفة $\mathbf{y}$ ببعد $m \times 1$ فإن الجداء الخارجي للمصفوفتين $\mathbf{x}$ و$\mathbf{y}$ هو مصفوفة A ببعد $n \times m$ حيث العنصر a_{ij} سيكون مساوياً $a_{ij}=x_{i1}y_{1j}$. وهذا أساسي جداً

لمعرفة كيفية تشكيل مصفوفة الارتباط من متحولين عشوائيين موزعين توزيعاً مشتركاً.

3.2.3 المعكوس الوهمي Pseudo-inverse

من المؤكد أن ليس لكل المصفوفات المربعة معكوس، ولكن كل مصفوفة لها معكوس وهمي، وهو عبارة عن نهاية لشكل معدل عن المصفوفة الأصلية. يوفر المقلوب الوهمي تقريباً بديلاً لحل نظام المعادلات المتزامنة (اللحظية). وتزداد فعالية المعكوس الوهمي في مسائل جعل مجموع المربعات أصغرياً.

إن مجموع المربعات الواجب جعله أصغرياً، فيما يخص الشبكات العصبونية الصنعية، هو عادة الأخطاء في خرج الشبكة. تمثل هذه الأخطاء الفرق بين الخرج المنشود **t** (target) المثالي والخرج الفعلي المحسوب **y** من خلال الشبكة عبر كل نماذج التدريب$\{(\mathbf{x}^p,\mathbf{t}^p)\ ,\ p = 1, 2 ,\dots ,P\}$، حيث $\mathbf{x}^p$ هو نموذج دخل التدريب رقم **p** و$\mathbf{t}^p$ هو الخرج المنشود الموافق له. ستكون الأشعة **x** ببعد n في حين تكون الأشعة **y** و**t** ببعد m، وسيكون الخطأ الكلي كما يلي:

$$E = \sum_{p=1}^{P} E^p = \sum_{p=1}^{p}\sum_{j=1}^{m}(t_j^p - y_j^p)^2 \qquad (15.3)$$

حيث E^p هو مجموع مربع الأخطاء من أجل كل نموذج **p** عبر كل وحدات الخرج، وسيكون شعاع الخرج $\mathbf{y}^p$ تابعاً للأوزان. مثلاً، إذا كانت المركبة j من الأشعة $\mathbf{y}^p$ هي تابعاً خطياً للأوزان :

$$y_j^p = \sum_{i=1}^{n} w_{ij} x_i^p$$

فإن تقليل الخطأ الكلي E للأوزان يتطلب أن تكون المشتقات الجزئية للمعادلة (15.3) مساوية للصفر كالآتــــي:

$$\begin{aligned}\frac{\partial E}{\partial w_{ij}} &= \frac{\partial}{\partial w_{ij}}\left(\sum_{p=1}^{P}\sum_{j=1}^{m}(t_j^p - \sum_{i=1}^{n} w_{ij} x_i^p)^2\right) \\ &= -2\sum_{p=1}^{P}\left(\sum_{i=1}^{n} t_j^p - w_{ij} x_i^p\right).x_i^p = 0\end{aligned} \qquad (16.3)$$

يمكن كتابة هذه العلاقة بالشكل المصفوفي كما يلي:

$$\mathbf{WXX}^T = \mathbf{TX}^T \Rightarrow \mathbf{WX} = \mathbf{T} \qquad (17.3)$$

حيث **W** مصفوفة الأوزان ببعد $m \times n$ ومركباتها w_{ij}، و**X** مصفوفة نماذج الدخـــل ببعد

$n \times p$ ومركباتها x_i^p، وT مصفوفة الخرج المنشود ببعد $m \times p$ ومركباتها t_j^p .

بوجه عام، لا يمكن أن يكون للمعادلة (3-17) حل مباشر باعتبار أن الجداء $\mathbf{XX^T}$ قد لا يكون له معكوس. وتحصل هذه الحالة عندما لا يملك الجداء $\mathbf{XX^T}$ m سطراً مستقلاً خطياً. لذلك، وبسبب عدم عكوسية $\mathbf{XX^T}$، فإن المعادلة تملك عدداً من الحلول.

لتحديد حلٍّ وحيد ينبغي فرض شرط إضافي على تعبير الأخطاء لتكون بقيمة أصغرية. وهذا الشرط المطلوب هو تحديد الأوزان تزامنياً ليكون الخطأ أصغرياً مع إضافة حد آخر يحتوي على مجموع مربعات كل الأوزان كما يلي:

$$E + \lambda \sum_{i=1} w_{ij}^2 \qquad (18.3)$$

حيث λ ثابت موجب. يقود جعل هذا التعبير أصغرياً إلى حل مصفوفي من الشكل التالي:

$$\mathbf{W(XX^T + \lambda I) = TX^T} \qquad (19.3)$$

من الممكن إثبات أنه من أجل كل $\lambda > 0$ فإن المصفوفة $(\mathbf{XX^T}+\lambda\,\mathbf{I})$ سيكون لها معكوس. لذا نضرب طرفي العلاقة (3-19) بالمعكوس $(\mathbf{XX^T}+\lambda\,\mathbf{I})^{-1}$ ونأخذ نهاية التعبير الناتج عندما $\lambda \to 0$، فنحصل على الحل المطلوب:

$$\mathbf{W} = \lim_{\lambda \to 0} [\mathbf{T\,X^T\,(XX^T} + \lambda\,\mathbf{I})^{-1}] = \mathbf{T}\,\mathbf{Ж} \qquad (20.3)$$

حيث $\mathbf{Ж}$ هو المعكوس الوهمي للمصفوفة $\mathbf{X}$، وسيكون هناك دائماً على الأقل حل واحد.

إذا وجدت حلول كثيرة، فالحل المتعلق بالمعكوس الوهمي الأول هو الحل ذو المجموع الأصغر لمربعـــات قيم المصفوفة. علاوة على ذلك، عندما يكون للمصفوفة $\mathbf{X}$ معكوس هو $\mathbf{X^{-1}}$ فإن $\mathbf{X^{-1}} = \mathbf{Ж}$.

عموماً، يعطى تعريف المعكوس الوهمي لمصفوفة مستطيلة $\mathbf{X}$ برتبة k. إن المعكوس الوهمي للمصفوفة $\mathbf{X}$ يحقق الشروط التالية:

$$1-\ \mathbf{X\,Ж X = X}$$
$$2-\ \mathbf{Ж\,X\,Ж = Ж} \qquad (21.3)$$

$3-\ \mathbf{X\,Ж}$ و $\mathbf{ЖX}$ مصفوفات هرميتية

المصفوفة الهرميتية (Hermitian) هي المصفوفة التـــي تساوي المرافق العقدي لمنقولها.

بالطبع، من أجل مصفوفة حقيقية فإن شرط هرميتيان يعنـــي أن المصفوفة متناظرة.

4.2.3 الأشعة الخاصة Eigenvectors

إذا كانت لدينا مصفوفة مربعة **A** ببعد $n \times n$ و **x** شعاع غير صفري ببعد n عنصر، فإن الشـــعاع **x** يسمى الشـــعاع الخاص للمصفوفة **A** إذا وُجد عـــدد مثل λ، بحيث يـــكون $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}$.

إن الأشعة الخاصة لمصفوفة معطاة هي الأشعة التـــي إذا ضربت بالمصفوفة حصلنا على نفس الشعاع مضروباً بثابت ما وليكن λ. يسمى الثابت λ بالقيمة الخاصة للمصفوفة **A** ذات الشعاع الخاص **x**. إن أي مصفوفة مربعة ببعد $n \times n$ لها n شعاعاً خاصاً وقيمة خاصة مرافقة. يمكن أن تكون بعض القيم الخاصة متساوية، وقد يكون بعضها أصفاراً والبعض الآخر عقدياً، وذلك حسب قيم عناصر المصفوفة **A**. إذا كان **x** شعاعاً خاصاً للمصفوفة **A**، فإن $k\mathbf{x}$ سيكون أيضاً شعاعاً خاصاً. وهكذا فإن الأشعة الخاصة تكون وحيدة مع ثابت الجداء فقط.

مثلاً، الشعاع الخاص **x** وقيمته الخاصة $\lambda = 6$، من أجل المصفوفة **A**، كما يلي:

$$\begin{bmatrix} 6 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 7 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ -12 \end{bmatrix}$$

$$\mathbf{A} \quad \mathbf{x} = \lambda\,\mathbf{x} \qquad (22.3)$$

هناك شعاعان خاصان آخران للمصفوفة **A** هما $(1 -1\ 0)$ و$(1\ 1\ 1)$، والقيم الخاصة يمكن إيجادها بسهولة. يمكن أن تكتب المعادلة $\mathbf{A}\,\mathbf{x} = \lambda\,\mathbf{x}$ بالشكل التالي:

$$(\mathbf{A} - \lambda\,\mathbf{I})\,\mathbf{x} = 0 \qquad (23.3)$$

وتسمى بالمعادلة المميزة للمصفوفة.

طبعاً لإيجاد القيم الخاصة للمصفوفة **A** علينا إيجاد n جذراً للمعادلة المميزة فيمـــا يتعلق بـــ λ وذلك باستعمال معينات المصفوفة **A**. سنهمل التفاصيل هنا، بغية الاختصار، وننصح القارئ المهتم بمراجعة أحد مراجع الجبر الخطي والفراغ الشعاعي.

أحياناً تسمى الأشعة الخاصة والقيم الخاصة بالأشعة المميزة والقيم المميزة، على الترتيب. وكما أشرنا من قبل، فإنها تؤدي دوراً هاماً في تطبيقات عديدة باعتبار أنها تكون مرتبطة

وسطاء الأنظمة الفيزيائية.

5.2.3 تحليل المركبة الأساسية Principal component analysis (PCA)

إن استعمال عدد أو عدة أعداد لوصف أو تمييز شيء أو مجموعة من الأشياء هو عملٌ شائعٌ في حقول عديدة. مثلاً، نستعمل علامة الامتحان للدلالة على مستوى معرفة الطالب المكتسبة من المنهاج المدروس، ونستخدم المتوسط والتباين لتلخيص مميزات توزع السكان، ...الخ.

إن تحليل المركبة الأساسية هو طريقة استُعملت للدلالة على خواص مجموعة من نماذج المعطيات متعددة المتغيرات. ويعتبر هذا التحليل طريقة تحويل خطي شائعة الاستعمال لتحليل المعطيات (مثلاً، استنباط المعالم أو الأشكال) أو دمج المعطيات. والتحويل هو عملية جعل التباين أعظمياً وهو بدوره يجعل مصفوفة التباين المشترك (covariance) لتوزيع نماذج الدخل مصفوفة قطرية.

إذا كانت لدينا أشكال دخل معطاة في فراغ شعاعي ببعد n، وأردنا إيجاد مجموعة جزئية ما ببعد m حيث $m < n$ من مجموعة الأشياء الكلية، تعبّر عن أكبر تباين ممكن للمعطيات، فإننا نقوم بإسقاط الفراغ ذي البعد n على الفراغ ذي البعد m، ومن ثم نختار كل مركبة في اتجاه التباين الأعظمي، وستكون هذه المركبات متعامدة مثنى.

إذا كانت لدينا P عينة لـــ n متغيراً، فإن هذه العينات ستكون أسطراً في المصفوفة X (التـــي تمثل مداخل الشبكة لـــ P نموذج تدريب، وسنوضح ذلك عند دراسة الشبكات) التالية:

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ x_{p1} & x_{p2} & \cdots & x_{pn} \end{bmatrix}$$

سنكوّن الآن مجموعة من n متحولاً جديداً y_j، حيث $j = 1,2,..,n$ ، هي "المركبات الأساسية"، من x_j التـــي هي تراكيب خطية من x_j . إذا كان W_i و y_i أشـــعة عمود ببعد n و p على الترتيب، يمكن عندها تلخيص طريقة PCA كالآتـــي:

1. لنأخذ المركبة الأساسية الأولى على طول اتجاه التباين الأعظمي، وليكن

$$\mathbf{y}_1 = \mathbf{X}\,\mathbf{w}_1 \qquad (24.3)$$

حيث $\mathbf{y}_1$ و$\mathbf{w}_1$ أشعة عمود. لاحظ أن هذه الكمية يجب أن تُقيَّد (تثبت)، لأن التباين سيكون أكبر فأكبر تماماً بزيادة قيم $\mathbf{w}_1$. طبعاً يمكن أن ينجز هذا التقييد بجعل كل أشعة الأوزان بطول مساوٍ للواحد، أي:

$$\mathbf{w}_1^T \mathbf{w}_1 = \sum_i \mathbf{w}_{i1}^2 = 1 \qquad (25.3)$$

2. سنجعل الآن مجموع المربعات أعظمياً وفقاً لقيمة $\mathbf{y}_1$، وذلك بتحقيق المساواة التالية:

$$\mathbf{y}_1^T \mathbf{y}_1 = \mathbf{w}_1^T \mathbf{X}^T \mathbf{X} \mathbf{w}_1 \qquad (26.3)$$

باستعمال طريقة لاغرانج يمكن حل مسائل التعظيم (Maximization)، من هذا النوع، مع تقييد متساوٍ على الأغلب.

إن تابع التركيب الجديد باستعمال المعادلتين (24.3) و(25.3) سيكون كما يلي:

$$L = \mathbf{w}_1^T \mathbf{X}^T \mathbf{X} \mathbf{w}_1 - \lambda_1 (\mathbf{w}_1^T \mathbf{w}_1 - 1) \qquad (27.3)$$

حيث λ_j هو ضارب لاغرانج. وبتحقيق الشرط (25.3) لكل قيم $\mathbf{w}_1$ يكون الحد الثاني مساوياً للصفر، ويأخذ التابع L شكل التابع الأصلي (26.3). للحصول على الحل المقابل لأكبر قيمة لـ L نأخذ المشتقات الجزئية لـ L لكل قيم $\mathbf{w}_1$ و λ_1، ونضع النتائج مساوية للصفر، فيكون لدينا:

$$\frac{\partial L}{\partial \mathbf{w}_1} = 2\mathbf{X}^T \mathbf{X} \mathbf{w}_1 - 2\lambda_1 \mathbf{w}_1 = 0$$
$$\mathbf{X}^T \mathbf{X} \mathbf{w}_1 = \lambda_1 \mathbf{w}_1 \qquad (28.3)$$

من هذه النتيجة والمعادلة (26.3) نستطيع كتابة:

$$\mathbf{y}_1^T \mathbf{y}_1 = \mathbf{w}_1^T \lambda_1 \mathbf{w}_1 = \lambda_1 \mathbf{w}_1^T \mathbf{w}_1 = \lambda_1 \qquad (29.3)$$

الحل $\mathbf{y}_1$ هو المركبة الأساسية الأولى مع تباين أعظمي λ_1. نلاحظ أيضاً أن λ_1 هي قيمة خاصة للجداء $\mathbf{X}^T\mathbf{X}$.

3. لإيجاد المركبة الثانية $\mathbf{y}_2$، نقوم بنفس الإجراءات المتبعة في $\mathbf{y}_1$ لكن نأخذ $\mathbf{y}_2$ بدلاً من $\mathbf{y}_1$. وهكذا يتحقق التعظيم بالمعادلة:

$$\mathbf{y}_2^T \mathbf{y}_2 = \mathbf{w}_2^T \mathbf{X}^T \mathbf{X} \mathbf{w}_2 \quad (30.3)$$

والمسألة تؤول إلى تحقيق الشرطين:

$$\mathbf{w}_2^T \mathbf{w}_2 = 1 \quad , \quad \mathbf{w}_1^T \mathbf{w}_2 = 0 \quad (31.3)$$

وإذا لاحظنا شرط التعامد الجديد، فسيكون لدينا:

$$L = \mathbf{w}_2^T \mathbf{X}^T \mathbf{X} \mathbf{w}_2 - \lambda_2 (\mathbf{w}_2^T \mathbf{w}_2 - 1) - \mu \mathbf{w}_1^T \mathbf{w}_2 \quad (32.3)$$

حيث λ_2 و μ هما مضاريب لاغرانج. وباتباع نفس الإجراءات السابقة، وبأخذ مشتقات جزئية لـ L بالنسبة إلى $\mathbf{w}_2$ ووضع النتيجة مساوية للصفر، نحصل على:

$$\mu = 2\mathbf{w}_1^T \mathbf{X}^T \mathbf{X} \mathbf{w}_2 = 2 \times 0 = 0$$
$$\mathbf{X}^T \mathbf{X} \mathbf{w}_2 = \lambda_2 \mathbf{w}_2 \quad (33.3)$$

نختار الآن قيمة λ_2 لتكون القيمة الخاصة الأكبر الثانية لـ $\mathbf{X}^T\mathbf{X}$.

4. بمتابعة العملية، نحصل على P قيمة خاصة $\lambda_1, \lambda_2, \ldots, \lambda_P$ ومصفوفة التعامد المرافقة $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_P]$. ويمكننا الآن الحصول على P مركبة أساسية للمصفوفة $\mathbf{Y}$ من الجداء $\mathbf{Y} = \mathbf{XW}$، حيث

$$\mathbf{Y}^T\mathbf{Y} = \mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W} = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ . & . & . & . \\ 0 & 0 & \ldots & \lambda_P \end{bmatrix} \quad (34.3)$$

باعتبار Λ هو قطر المصفوفة مع باقي العناصر أصفار. نستنتج أن المركبات الأساسية للمصفوفات$\mathbf{Y}$ تكون غير مترابطة (متعامدة) مثنى، ومجموع المربعات سيكون λ_i .

عندما تكون رتبة المصفوفة $\mathbf{X}$ هي s، حيث $s < P$، فإن القيم الخاصة P-s ستكون أصفاراً. في هذه الحالة، نستعمل المتحولات المستقلة التـــي عددها s للتعبير عن تباين عينات المعطيات (يصحّ أيضاً أن تكون بعض القيم الخاصة صغيرة جداً). عندئذ يعطى التغير الكلي في $\mathbf{X}$ (أخذت المجاميع السفلية عبر n عينة مراقبة):

$$\sum \mathbf{x}_1^2 + \sum \mathbf{x}_2^2 + \cdots + \sum \mathbf{x}_p^2 = Tr(\mathbf{X}^T\mathbf{X}) = Tr(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}) = \sum_{j=1}^{p} \lambda_j = \sum_{j=1}^{p} \mathbf{Y}_j^T\mathbf{Y}_j \qquad (35.3)$$

باعتبار $\mathrm{Tr}(\mathbf{W}^T\mathbf{X}^T\mathbf{X}\mathbf{W}) = \mathrm{Tr}(\mathbf{X}^T\mathbf{X}\mathbf{W}\mathbf{W}^T) = \mathrm{Tr}(\mathbf{X}^T\mathbf{X})\ ,\ \mathbf{W}\mathbf{W}^T = \mathbf{I}$

تعريف: أثر المصفوفة المربعة **A** ذات البعد $n \times n$ يعطى كما يلي:

$$\mathrm{Tr}\ \mathbf{A} = \sum_{i=1}^{n} a_{ii}$$

ويمكن استنتاج الخواص التالية من هذا التعريف :

1- $\mathrm{Tr}(x_1\mathbf{A} + x_2\mathbf{B}) = x_1\ \mathrm{Tr}\ \mathbf{A} + x_2\ \mathrm{Tr}\ \mathbf{B}$ (x_1, x_2 قيم سُلّمية)

2- $\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA})$ ($n \times m$ رتبة المصفوفة **B**، $m \times n$ رتبة المصفوفة **A**)

3- $\mathrm{Tr}\ \mathbf{A} = \sum_{i=1}^{n} \lambda_i$ (λ_i القيم الخاصة للمصفوفة **A**)

يمكن للمرء أن يثبت أن عوامل ارتباط العزوم ρ_{ij} الناتجة بين $\mathbf{x}_i$ و$\mathbf{y}_j$ تحقق العلاقة التالية :

$$\frac{w_{i1}^2}{\sum \mathbf{x}_j^2}\lambda_1 + \frac{w_{i2}^2}{\sum \mathbf{x}_j^2}\lambda_2 + \cdots + \frac{w_{iP}^2}{\sum \mathbf{x}_j^2}\lambda_P = \rho_{i1}^2 + \rho_{i2}^2 + \cdots + \rho_{iP}^2 = 1$$

في الفصل الثالث عشر سنعطي مثالاً عن فائدة PCA لنوع خاص من خوارزميات تعليم الشبكة العصبونية الصنعية.

3.3 مراجعة لبعض مفاهيم الحساب

نفترض أن القارئ يعرف جيداً طرق حساب النهايات والتوابع المستمرة. إن نهاية التابع $f(x)$ عندما تنتهي x إلى القيمة a يرمز لها بالشكل: $\lim_{x\to a} f(x) = L$ ، والنهاية إن وجدت، تساوي L.

1.3.3 التفاضل

ليكن $\Delta x = x_2 - x_1$ هو الفرق بين النقطتين x_1 وx_2، وليكن لدينا التابع: $y = f(x)$، عندها يكون مشتق التابع $f(x)$ بالنسبة إلى المتحول x (ويرمز له $dy / dx = Dy$) هو:

$$Dy = dy/dx = \lim_{\Delta x \to 0} \frac{f(x+\Delta x)}{\Delta x} \tag{39.3}$$

وهذا المشتق موجود إذا كانت النهاية موجودة. والمشتق هو مقدار لحظي لتغير التابع عند كل نقطة للمتحول المستقل x.

يستطيع المرء دائماً إيجاد المشتق، إذا كان موجوداً، من التعريف السابق. مثلاً، يوجد مشتق كثير الحدود $f(x) = -2x^2 + 3x - 10$ مباشرة كما يلي:

$$\frac{f(x+\Delta x) - f(x)}{\Delta x} = \frac{[-2(x+\Delta x)^2 + 3(x+\Delta x) - 10] - (-2x^2 + 3x - 10)}{\Delta x}$$

$$= \frac{-4x\Delta x - 2(\Delta x)^2 + 3\Delta x}{\Delta x} = -4x - 2\Delta x + 3$$

ومن ثم فإن مشتق التابع $f(x)$ هو تماماً:

$$Dy = dy/dx = \lim_{\Delta x \to 0} (-4x - 2\Delta x + 3) = -4x + 3$$

يرمز لمشتق التابع $f(x)$ أيضاً بـــ ((القيمة البدائية لـــ x) f – $f'(x)$). إن أبسط طريقة لإيجاد مشتق التابع هي باستعمال قواعد الاشتقاق من أجل أشكال تابعية مختلفة.

وبطريقة مشابهة يمكن حساب مشتقات المرتبة الأعلى. فمثلاً، المشتق الثانـــي للتابع $f(x)$ هو مشتق المشتق الأول لهذا التابع. وهكذا فإن المشتق الثانـــي للتابع في المثال السابق $f(x) = -2x^2 + 3x - 10$ سيكون مشتق التابع $-4x + 3$، أو

$$D^2y = d^2y/dx^2 = \lim_{\Delta x \to 0} \frac{[-4(x+\Delta x) + 3] - (-4x + 3)}{\Delta x} = \frac{-4\Delta x}{\Delta x} = -4$$

المشتق الثانـــي هو مقدار لحظي لتغير المشتق عند كل نقطة للمتحول المستقل x. وتعّرف جميع مشـــتقات المرتبة الأعلى بنفـــس الطريقة. إذا كان $y = f(u)$ تابعاً قابلاً للاشـــتقاق و $u = g(x)$ قابلاً للاشتقاق أيضاً، فإن مشتق التابع y بالنسبة إلى x يمكن إيجاده بتطبيق قاعدة السلسلة (chain rule) كما يلي:

$$\frac{dy}{dx} = \frac{dy}{du}\frac{du}{dx} \tag{37.3}$$

مثلاً، ليكـــن $y = f(u) = u^3 - 6u$، حيـــث $u = g(x) = x^4 + 6x$ فـــإن

$dy/dx = (3u^2 - 6)(4x^3 + 6)$. بالطبع، يمكن تعميم المفاهيم السابقة على توابع بمتحولات متعددة، حيث تعرّف المشتقات الجزئية بطريقة مشابهة، مشتق أول جزئي لكل متحول وتعتبر باقي المتحولات قيم ثابتة. مثلاً، ليكن $f(x, y)$ تابعاً لمتحولين مستقلين هما x وy، فإن المشتقات الجزئية للتابع f بالنسبة إلى كل متحول تعطى كما يلي:

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \quad , \quad \frac{\partial f}{\partial y} = \lim_{\Delta x \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

بالطبع، يصحّ هذا كله إذا وُجدت النهايات.

تمثل كل من هذه المشتقات مقداراً لحظياً لتغير التابع بالنسبة للتغيرات في كل من المتحولات المستقلة على نحوٍ منفصل. مثلاً، ليكن $f(x, y) = 3x^3 + 6xy^4 + 2y$، يعطى المشتق الجزئي للتابع f بالنسبة للمتحول x باشتقاق التابع f واعتبار المتحول y كثابت، وكذلك مشتق التابع f بالنسبة للمتحول y يوجد بطريقة مشابهة (حيث نعتبر x ثابتاً) كما يلي:

$$\partial f/\partial x = 9x^2 + 6y^4 \quad , \quad \partial f/\partial y = 24xy^3 + 2$$

تُستعمل الحسابات الجزئية كثيراً في مسائل الاستمثال، حيث يُطلب إيجاد القيمة العظمى والصغرى للتابع.

وتُستعمل بوجه خاص في أسس تنفيذ التعليم لعدة شبكات عصبونية على معيار تابع الخطأ التربيعي الأصغري. وهناك أمثلة عديدة دُرست لجعل تابع الخطأ أصغرياً، قد ذكرت في هذا الكتاب.

2.3.3 التكامل

استُعمل حساب التكامل لإيجاد المساحة المحصورة بين منحنيات وحدود أخرى محيطة بهذه المساحة. أيضاً، إذا كان مشتق التابع الأصلي معروفاً، والتابع الأصلي غير معروف، فإن حساب التكامل يعتبر طريقة لإيجاد التابع الأصلي.

ليكن f تابعاً معطى، يمكن استعمال الطريقة المذكورة في المقطع السابق لإيجاد مشتق هذا التابع. من ناحية أخرى، إذا أعطينا المشتق f'، قد يكون من المأمول معرفة التابع الأصلي f

لهذا المشتق، والذي يسمى عكس المشتق (antiderivative).

تسمح لنا معرفة مشتقات توابع مختلفة بإيجاد مجموعات من التوابع، أي توابع عكس المشتق. وهي توابع، ليست وحيدة عموماً باعتبار أن مشتق الثابت يساوي الصفر. لذا يجب فرض قيود أخرى لإيجاد توابع عكس المشتق خاصة.

إن عملية إيجاد تابع عكس المشتق هي عملية تكامل، ومجموعة التوابع الموجودة بواسطة هذه العملية تدعى بالتكامل غير المحدود (indefinite integral)، يكتب التكامل غير المحدود للتابع $f(x)$ كما يلي: $\int f(x)dx$. مثلاً، التكامل غير المحدود للتابع $f(x) = 2x + 6$ x^2+6x+c، حيث C هو ثابت التكامل، وهو قيمة غير معروفة. للتأكد من أن x^2+6x+c هو تكامل التابع $f(x)$ ، يستطيع المرء أخذ مشتق x^2+6x+c .

إذا كان f تابعاً مستمراً، نعرف التكامل غير المحدود لهذا التابع كما يلي: $\int f(x)dx = F(x)+C$ ، حيث $F'(x) = f(x)$. عملياً، يمكننا استعمال مجموعة من القواعد أو الجداول لإيجاد تكامل التوابع. بالطبع هذه القواعد تكون مرتبطة مباشرة بمشتقات التكامل.

إن معرفة التكامل مطلوبة لحل الكثير من المسائل الحقيقية. توفر معرفة تكامل منحنٍ ما حساب المساحة أسفل هذا المنحنـــي، وأيضاً يمكن أن تفسر كنهاية. وبوجه خاص، إذا كان f تابعاً محدوداً ضمن المجال $[a ,b]$، فإن تكامل التابع على هذا المجال يعرف كما يلي:

$$\int_a^b f(x)dx = \lim_{n\to\infty}\sum_{i=1}^{n} f(x_i)\Delta x_i = A \qquad (38.3)$$

حيث $\Delta x_i = \frac{b-a}{n}$.

بالطبع يجب ألاّ ننسى أن هذا التكامل معرّف إذا وجدت النهاية. لاحظ كلما كانت قيمة الانتقال Δx_i في التقسيم الجزئي للمجال $[a, b]$ تقترب من الصفر فإن عدد الانتقالات n يقترب من اللانهاية ($n \to \infty$). القيمتان a و b هما الحد الأدنى والأعلى للتكامل على الترتيب.

تنص النظرية الأساسية لحساب التكامل على أنه إذا كان F عكس المشتق للتابع المستمر f عبر مجال ما، عندئذ مهما تكن $x = a$ و $x = b$، حيث $a \leq b$، فإن

$$\int_a^b f(x)dx = F(b) - F(a)$$

لاحظ أن ثابت التكامل حذف من العلاقة السابقة. مثلاً، لإعطاء قيمة لتكامل كثير الحدود $2x^2 + 4x - 5$ على المجال [1,4]، لدينا

$$\int_1^4 (2x^2 + 4x - 5)dx = F(4) - F(1) = \left[\frac{2x^3}{3} + 2x^2 - 5x\right]_1^4$$

حيث يشير الرمز $]_1^4$ إلى حدود التكامل. وهكذا سيكون

$$F(4) - F(1) = \left[\frac{2(4)^3}{3} + 2(4)^2 - 5(4)\right] - \left[\frac{2(1)^3}{3} + 2(1)^2 - 5(1)\right] = \frac{171}{3}$$

وكما في الاشتقاق، هناك قواعد عديدة للتكامل تساعد على تقييم تكاملات محدودة وغير محدودة.

يوجد أيضاً تكاملات متعددة بواسطة أخذ عكس مشتقات متعددة عبر مجالات منفصلة للتكامل. بالمثل، تعرف تكاملات التوابع المتعددة المتحولات لكل متحول منفصلاً عن البقية، وهذا نوعاً ما شكل معاكس للمشتقات الجزئية. سنهمل التفاصيل هنا ونقترح على القارئ المهتم بمراجعة أحد الكتب في الحساب التفاضلي والتكاملي.

3.3.3 المعادلات التفاضلية

المعادلات التفاضلية هي المعادلات التـــي تتضمن حدوداً تفاضلية أو اشتقاقية. استُعملت المعادلات التفاضلية لوصف سلوك الأنظمة التغيرية بما فيها الشبكات العصبونية الديناميكية. تنفذ المعادلات التفاضلية العادية ODEs (Ordinary Differential Equations) مشتقات لتابع بمتحول مستقل واحد، مثل الزمن. ورتبة المعادلة التفاضلية هي المشتق ذو المرتبة العليا الموجود في المعادلة، أما درجة المعادلة فهي قوة المشتق ذات المرتبة العليا في المعادلة.

لكي نحل المعادلة التفاضلية علينا إيجاد توابع عكس المشتقات لكل حدود المشتقات الموجودة في المعادلة. مثلاً، نظام المعادلات التفاضلية أحادية المتحول المستقل المستعملة لوصف سلوك الشبكة العصبونية الديناميكية هو كما يلي:

$$\frac{dy_j}{dt} = \sum_i w_{ij}x_i + \sum_{k\in s_j} v_{kj}y_k - h(y_j)$$
$$\frac{dw_{ij}}{dt} = \alpha(x_i - w_{ij}) \qquad (39.3)$$

والمعادلة الثانية تعرف ديناميكيات التعليم للشبكة، حيث المعدل الزمنـــي لتغير الأوزان هو تابع للمداخل والأوزان. سنعطى تفاصيل أكثر عن هذه الأنظمة في مقاطع لاحقة تصف أداء شبكات عصبونية خاصة.

4.3 مراجعة لمفاهيم الاحتمالات

المتحول العشوائي r.v (random variable)، X هو تابع معرف على فراغ العينة Ω؛ أي فراغ النتائج الممكنة للتجربة، سنستعمل الأحرف الكبيرة لرموز المتحولات العشوائية، والأحرف الصغيرة لقيم المتحول العشوائي الممكن فرضها.

يدعى التابع الذي يحدد احتمال كل نتيجة ممكنة للتجربة بتابع كثافة الاحتمال pdf (probability density function) للمتحولات المستمرة، وبتابع كتلة الاحتمال pmf (probability mass function) للمتحولات المتقطعة.

يجب أن يحقق المتحول العشوائي المستمر X، الذي له تابع كثافه $f(x)$ الشروط التالية:

1- $f(x) \geq 0$ لكل قيم x

2- $\int_{-\infty}^{\infty} f(x)dx = 1$

يعنـــي الشرط الأول أن الاحتمالات ليست صفرية، ويؤكد الشرط الثانـــي أن الحوادث تكون متنافية بالتبادل ومتضادة (متامة) تماماً (mutually exclusive and collectively exhaustive). وبالمثل تكون نفس الشروط محققة من أجل المتحولات العشوائية المتقطعة.

يمكن التعبير عن الشرطين السابقين رياضياً كما يلي:

$$A \cap B = \Phi \Rightarrow P(A \cup B) = P(A) + P(B)$$
$$A_i \cap A_j = \Phi, i \neq j \Rightarrow P(\bigcup_i A_i) = \sum_i P(A_i)$$

أو $A_i \supset A_{i+1}, \bigcap_i A_i = \Phi \Rightarrow \lim_{n\to\infty} P(A_n) = 0$

حيث B وA أحداث، وهي عبارة عن مجموعات جزئية من فراغ العينة Ω.

إن احتمال أن يكون للمتحول العشوائي المستمر X قيمة في المجال بين a وb حيث $a < b$، يساوي المساحة المحصورة تحت تابع الكثافة بين النقطتين $x = a$ و$x = b$، ويعطى ذلك بتكامل محدود للتابع $f(x)$ بين $x = a$ و$x = b$، يرمز له بواسطة $P(a \le X \le b)$، أي:

$$P(a \le X \le b) = \int_a^b f(x)dx \qquad (40.3)$$

تعطى قوانين الاحتمالات باختصار كما يلي:

$$\begin{aligned}
&P(A^c) = 1 - P(A) \\
&P(A \cup B) = P(A) + P(B) - P(A \cap B) \\
&P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\
&\qquad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\
&P(B \setminus A) = P(B) - P(A \cap B) \\
&P(A \Delta B) = P(A) + P(B) - 2P(A \cap B)
\end{aligned} \qquad (41.3)$$

حيث B وA أحداث، وهي عبارة عن مجموعات جزئية من فراغ العينة Ω، حيث يشير A^c لعملية المتمم، ولعملية الفرق، وΔ لعملية الفرق التناظري.

يعرف تابع التوزيع التجميعي (cumulative) لمتحول عشوائي مستمر X ذي تابع كثافة $f(x)$، يرمز له $F(x)$، كما يلي:

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx \qquad (42.3)$$

من المفيد معرفة بعض وسطاء التوزيعات الاحتمالية عند دراسة سلوك الحوادث أو الأنظمة الموصوفة بالاحتمال أو الصدفة، مثل المتوسط (mean)، والتباين (variance)، والوسط (median)، والمجال (range)، والنمط (mode)، والنسب المئوية (percentiles)، والعزوم (moments).

الوسط هو قيمة المتغير التـــي تشير إلى مركز التوزيع مع نصف كتلة الاحتمال تحت قيمة الوسط ونصف الكتلة فوق القيمة.

النمط هو نقطة أو نقاط الاحتمال الأعظمي.

المجال هو الانتقال للمتغير الذي لا يكون توزيع الاحتمال عبره مساوياً للصفر.

النقاط المئوية هي القيم المتغيرة التـــي تكون فيها كتلة الاحتمال مساوية للقيمة المئوية. مثلاً، الوسط هو نقطة 50% والربع هو نقطة 25%.

عزوم التوزيع (إن وجدت) تميز التوزيع تمييزاً كاملاً. يعرف العزم رقم k ($k = 0, 1, 2, \ldots$) لتوزيع الاحتمال بالعلاقة التالية:

$$\mu_k = \int_a^b x^k f(x)dx \qquad (43.3)$$

بالطبع إذا كان العزم موجوداً. إن العزم الأول والثانـــي هما الأكثر استخداماً لإعطاء النقطة المركزية أو القيمة المتوسطة ونشر أو تباين التوزيع، على الترتيب.

يعرف **التباين** σ^2 كالعزم الثانـــي حول المتوسط، **والانحراف المعياري** σ هو الجذر التربيعي للتباين.

$$Var[X] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \qquad (44.3)$$

حيث $\mu \equiv \mu$.

ليكن لدينا تابع لمتحول عشوائي X ذي تابع كثافة احتمال $f(x)$، مثل $h(X)$، يعرف **المتوسط** أو **القيمة المتوقعة** للتابع كما يلي:

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx \qquad (45.3)$$

من هذا التعريف نرى أن $\mu_k = E[x^k]$.

يعتبر التوزيع بقيم مستمرة الأكثر شيوعاً هو التوزيع الطبيعي (normal)، حيث هو تابع أحادي المتغير ذو تابع كثافة احتمال من الشكل:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-1/2[(x-\mu)/\sigma]^2} \qquad (46.3)$$

مهما تكن $-\infty < x < \infty$.

المتوسط والتباين للتوزيع الطبيعي هما μ وσ^2 على الترتيب، والانحراف المعياري هو σ.

هناك توزيعات أخرى تحدث كثيراً، مثل ثنائي الحد (binomial)، والأســـي

(exponential)، والهندسي (geometric)، والمنتظم (uniform)، وبواسون (Poisson)، ...الخ. في هذه المراجعة المختصرة سنهمل التفاصيل.

يمكن أن تقدر وسطاء التوزيعات؛ مثل الشكل التابعي pdf والمتوسط والتباين ووسطاء أخرى، من العينات (نتائج المراقبة) المأخوذة من التوزيع، وقد تكون بعض العينات إحصائية (stochastic). باعتبار أن التابع للمتحول العشوائي هو متحول عشوائي أيضاً؛ فإن أي إحصاء هو متحول عشوائي. لتكن $x_1, x_2, ..., x_n$ عينة من n مراقبة مستقلة مأخوذة من نفس التوزيع، وموزعة توزيعاً متماثلاً ومستقلاً iid؛ (independent, identically distributed).

المقدر الأفضل لمتوسط المجتمع هو (من أجل عدة توزيعات) تماماً متوسط العينة (سنشرح لاحقاً ماذا نعنـــي بكلمة أفضل):

$$\bar{x} = \sum_{i=1}^{n} x_i$$

المقدرات الجيدة تتمتع ببعض الخواص التالية:

1. مقدر غير منحاز؛ هو الذي تساوي القيمة المتوسطة له الوسيط المقدر. متوسط العينة هو مقدر غير منحاز لمتوسط المجتمع، باعتبار أن:

$$E[\bar{x}] = E\left[\frac{1}{n}\sum_{i=1}^{n} x_i\right] = \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[x_i] = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu \qquad (78.3)$$

2. سيكون لتوزيع احتمال المقدر الجيد تلاش أصغري، هذا يعني، من بين كل المقدرات سيكون له التباين الأصغر. ويمكن إثبات، أن متوسط العينة يحقق بالفعل تبايناً أصغرياً لعدة توزيعات.

3. أخيراً، لما كان حجم العينة للإحصاء الجيد متزايداً، فيمكن أن يتوقع المرء اقتراب الإحصاء من قيمة الوسيط الصحيح. بكلمات أخرى، سيتقارب إلى وسيط المجتمع المقدر. تعرّف المقدرات المتميزة بهذه الخاصية بالإحصاءات الكافية.

يمكن أن تعمم المفاهيم المذكورة سابقاً لتشمل التوزيعات بعدة متحولات؛ أي التوزيعات بأكثر من متحول عشوائي واحد. في بعض التوزيعات، يمكن تعريف وسطاء أخرى تصف العلاقة بين المتحولات المختلفة.

ليكن $f(x, y)$ تابع كثافة احتمال مشترك للمتحولين العشوائيين X وY. تسمى التوزيعات

أحادية التغير للمتحول X، $f(x)$، وللمتحول Y، $f(y)$، بالتوزيعات الأحادية. يمكن أن توجد مثل هذه التوزيعات منَ التوزيع المشترك بواسطة تكامل كوشي (Cauchy)، على الشكل التالي:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy \quad , \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx \qquad (48.3)$$

وتسمى بالتوزيعات الهامشية (**Marginal distributions**).

تعطى القيمة المتوقعة للتابع g(X,Y) بالعلاقة التالية:

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)dxdy \qquad (49.3)$$

وتعطى متراجحة شوارتز بالشكل التالي:

$$(E[XY])^2 \leq E[X^2]E[Y^2] \qquad (50.3)$$

أحد الوسطاء الخاصة الهامة للتوزيعات المتعددة المتغيرات هو عامل الارتباط ρ.

يقيس عامل الارتباط علاقة المتغير مع متغير آخر؛ أي اعتماد المتحول الأول على الآخر. يعرف الارتباط بين متحولين عشوائيين كنسبة جداء التباينين المشتركين إلى جداء الجذر التربيعي لتباين كل من المتحولين :

$$\rho = \frac{cov(x)\,cov(y)}{\sqrt{var(x)}\,\sqrt{var(y)}} = \frac{E[(x-\mu_x)(y-\mu_y)]}{\sigma_x\ \sigma_y} \qquad (51.3)$$

حيث μ_x وμ_y هما متوسطات المتغيرات X وY على التالي و σ_x وσ_y هما الانحرافات المعيارية الموافقة، ويعرّف التباين المشترك للمتحولين العشوائيين X وY كما يلي:

$$\begin{gathered} Cov[X,Y] = E[(X-\mu_1)(Y-\mu_2)] \\ E[X] = \mu_1 \quad , \quad E[X] = \mu_2 \end{gathered} \qquad (52.3)$$

لاحظ أن $-1 \leq \rho \leq 1$. إذا كان X وY مستقلين فإن $\rho = 0$، والمتحولان في هذه الحالة يكونان غير مرتبطين. إذا تحرك المتحولان أحدهما مع الآخر فإن $\rho = 1$، وإذا تحرك أحدهما بعكس الآخر فإن $\rho = -1$.

إذا كان X وY متحولين عشوائيين مستقلين ، فيمكن تلخيص التعاريف التالية:

$$
\begin{aligned}
&f(x,y) = f_X(x)f_Y(y) \\
&F(x,y) = F_X(x)F_Y(y) \\
&E[XY] = E[X]E[Y] \\
&Var[X+Y] = Var[X] + Var[Y] + 2Cov[X,Y] \\
&Cov[X,Y] = 0
\end{aligned}
\tag{53.3}
$$

هناك احتمالات أخرى سنتعرض لها عند الحاجة.

5.3 مراجعة لمفاهيم نظرية المعلومات

نشأت نظرية المعلومات وتطورت بفضل أعمال Claude Shannon عام 1940 [13]. عرّف شانون قياساً كمياً للمعلومات مبنياً على كمية الشك المكتشفة في المعلومات المستقبلة حديثاً. وارتبط عمله الأصلي ببث المعطيات عبر قنوات الاتصالات، لكن هذا العمل سرعان ما وجد له تطبيقات واسعة في كثير من الحقول مثل الفيزياء، والإحصاء، والهندسة، وعلوم الحاسوب، وعلم النفس، ...الخ.

المبدأ الأساسي الكامن خلف نظرية المعلومات هو: إذا نشأ اتصال بين مرسل ومستقبل، فإن هذا الاتصال يوفر للمستقبل بعض الحقائق الجديدة حول حادثة ما، أي إنه يكتسب بعض المعلومات. من ناحية أخرى، توفر الحقائق الجديدة غير المحتملة (غير المعروفة؛ مفاجآت أو ربما حقائق غير عادية) معلومات كثيرة. هذا هو الأساس الذي اعتمده العالم Shannon لقياس المعلومات والذي أطلق عليه اسم الأنتروبـــي Entropy. إنه قياس لمقدار الشك (uncertainty) لمحتوى رسالة أو متحول عشوائي مفرد.

عُرِّف قياس شانون للمعلومات بكمية الشك السابق في ناتج تجربة عشوائية عندما جرت مراقبة هذا الناتج. في حالة مجموعة منتهية من الحوادث، إذا كان احتمال الحادثة X معطى بواسطة $p(x)$ (سنستعمل X للدلالة على المتحول العشوائي بنتائج محتملة x)، فإن المعلومات المحتواة في تلك الحادثة تعطى بالعلاقة التالية: $\log(1/p(x))$.

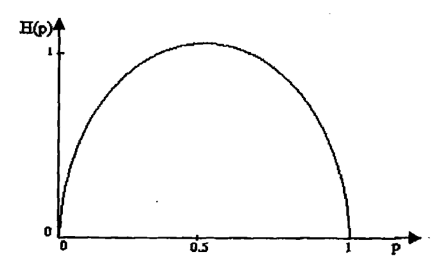
وهكذا، عندما يصبح الحدث مؤكداً أكثر، أي $p(x) \to 1$، فإن كمية المعلومات المحتواة في الحادثة تتضاءل إلى الصفر، وإذا كان احتمال الحدث صغيراً جداً $(p(x) \approx 0)$، فإن المعلومات المحتواة في الحدث تكون كبيرة. من أجل عدد محدد من الحوادث x_i، حيث

N, ..., $i = 1, 2$، يعرّف معدل المعلومات أو الأنتروبـي H(X) كما يلي :

$$H(X) = -\sum_{j=1}^{N} p(x_i)\log p(x_i) = -E(\log X) \qquad (54.3)$$

حيث E هي القيمة المتوقعة.

لاحظ أن H(X) يكون أعظمياً عندمـا تكـون الحوادث غير متسـاوية الاحتمال، أي $x_i = 1/N$ حيث N, ..., $i = 1, 2$، وأصغرياً عندما تكون الحوادث معروفة. هذا موضح في (الشكل 2.3) حيث يعطى مخطط H كتابع لـ p في حال $N = 2$ و$p(x_i) = p$ و$P(x_2) = 1 - p$.



الشكل 2.3: الأنتروبـي كتابع احتمال الحدث

من التعريف يمكننا استنتاج أن $H(X) \geq 0$. لاحظ أن H(X) يكون أعظمياً عند $p = 0.5$ ومن ثم ينحدر سريعاً إلى الصفر أو الواحد. مثلاً في عملية رمي مكعب سداسي الأوجه، سيكون الأنتروبـي تقريباً $H = \log_2(6) = 2.56\,\text{bits}$، وهذه هي المعلومات المكتسبة عند مراقبة النتائج. لاحظ أننا استخدمنا اللوغاريتم ذا الأساس 2 للحصول على بتات (bits) عوضاً عن استخدام اللوغاريتم الطبيعي للحصول على (nats). بالطبع يمكننا كتابة العلاقة التـي تحول أساس اللوغاريتم على الشكل التالي:

$$H_b(X) = (\log_b a)H_a(X)$$

مثال 1:

ليكن $X = \{a, b, c, d\}$ باحتمالات (1/2, 1/4, 1/8, 1/8) على الترتيب. يعطى الأنتروبـــي لـــ X بـــ:

$$H(X) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} - \frac{1}{8}\log\frac{1}{8} = \frac{7}{4}\text{bits}$$

يُعرَّف الأنتروبـــي المشترك لمتحولين عشوائيين أو أكثر بطريقة مشابهة، حيث يعطى الأنتروبـــي للمتحولين X وY الموزعين توزيعاً مشتركاً بالعلاقة التالية:

$$\begin{aligned} H(X,Y) &= -E(\log[p(X,Y)]) \\ &= -\sum_{x}\sum_{y} p(x,y)\log[p(x,y)] \end{aligned} \qquad (55.3)$$

يمكن إثبات أن الأنتروبـــي المشترك (joint entropy) لزوج من المتحولات لا يزيد مطلقاً عن مجموع الأنتروبيات مفردة (أي الأنتروبيات الهامشية)، هذا يعني أن $H(X,Y) \le H(X) + H(Y)$. تتحقق المساواة عندما يكون X وY متحولين مستقلين.

من توزيعات الاحتمال الشرطية، نستطيع تعريف الأنتروبـــي الشرطي لـــ Y مع X معطى، ويرمز له بـــ $H(Y|X)$، كما يلي:

$$\begin{aligned} H(Y|X) &= -E(\log[p(Y|X)]) = \sum_{x} p(x)H(Y|X=x) \\ &= -\sum_{x} p(x)\sum_{y} p(y|x)\log[p(y|x)] \qquad (56.3) \\ &= -\sum_{x}\sum_{y} p(x,y)\log[p(y|x)] \end{aligned}$$

حيث أخذت المجاميع عبر كل القيم الممكنة لـــ x وy . تقيس المعادلة (56-3) الشك المتبقي حول Y بعد معرفة X. من الواضح أنه أقل من الشـــك حول Y لوحدها (دون معرفة X).

يمكن أن نرى، باستعمال العلاقة بين توزيعات الاحتمال المشترك والشرطي، أن الأنتروبـــي المشترك هو مجموع الأنتروبيات الشرطية والهامشية وذلك بتطبيق قاعدة السلسلة:

$$H(X,Y)=H(X)+H(Y|X) \quad (57.3)$$

ويمكننا أيضاً من هذه القاعدة إنشاء العلاقة التالية:

$$H(X,Y|Z)=H(X|Z)+H(Y|X,Z) \quad (58.3)$$

والقاعدة يمكن أن توسع إلى أي عدد من المتحولات العشوائية.

مثال 2:

ليكن (X ,Y) متحولين عشوائيين لهما التوزيع المشترك التالي:

Y/X	1	2	3	4
1	1/8	1/16	1/32	1/32
2	1/16	1/8	1/32	1/32
3	1/16	1/16	1/16	1/16
4	1/4	0	0	0

التوزيع الهامشــي لــ X هو (1/2, 1/4, 1/8, 1/18)، والتوزيع الهامشــي لــ Y هو (1/4 ,1/4 ,1/4 ,1/4)، وبالتالي $H(X)=7/4$ bits و$H(Y)=2$ bits. أيضاً لدينا:

$$H(X|Y)=\sum_{i=1}^{4}p(Y=i)H(X|Y=i)$$

$$=1/4H(1/2\ ,1/4\ ,1/8\ ,1/18)$$

$$+\ 1/4H\ (1/4\ ,1/2\ ,1/8\ ,1/8)$$

$$+\ 1/4H(1/4\ ,1/4\ ,1/4\ ,1/4)$$

$$+\ 1/4H(1\ ,0\ ,0\ ,0)$$

$$=1/4\times7/4+1/4\times7/4+1/4\times2+1/4\times0$$

$$=11/8 \text{ bits}$$

وبالمثـــل نجـــد أن $H(Y|X)=13/8$ bits و$H(X,Y)=27/8$ bits. لاحـــظ أن $H(Y|X)\neq H(X|Y)$، ولكن تذكر أن $H(X)-H(X|Y)=H(Y)-H(Y|X)$.

إن القياس المفيد للمسافة بين توزيعي احتمال هو الأنتروبـــي النسبـــي (relative entropy) (يُعرف أيضاً بمسافة Kullback Leibler). لتكن p(x) وq(x) توزيعات الاحتمال للمتحولين العشوائيين X وY، يعرف الأنتروبـــي النسبـــي كما يلي:

$$D(p\|q) = E(\log[p(X)/q(X)]) \\ = \sum_x p(x)\log[p(x)/q(x)] \tag{59.3}$$

حيث بالتعريف، $0\log(0/q) = 0$ و $p * \log(p/0) = \infty$. الأنتروبـــي النسبـــي هو دائماً قيمة ليست سالبة وقد يكون صفراً إذا وفقط إذا $p = q$. إنه ليس مسافة مترية، باعتبار أنه ليس متناظراً ولا يحقق المتراجحة المثلثية. ومع ذلك، فهو مفيد للتعبير عن المسافة بين توزيعين.

قياس نظري آخر للمعلومات وله تفسير فيزيائي هو: المعلومات المتبادلة بين المتحولين العشوائين X وY. المعلومات المتبادلة (mutual information) هي قياس لكمية المعلومات التـــي يحويها متحول عشوائي أول حول متحول عشوائي آخر. تعتبر هذه العملية تخفيضاً لمقدار الشك في متحول عشوائي تبعاً لمعرفتنا لمتحول عشوائي آخر.

ليكن $p(x,y)$ تابع كتلة الاحتمال المشترك و$p(x)$ و$q(y)$ هما تابعا كتلة الاحتمال الهامشي للمتحولين العشوائيين X وY. يعطى تعريف المعلومات المتبادلة بين X وY بالمعادلة التالية:

$$I(X;Y) = E\left(\log\frac{p(X,Y)}{p(X)p(Y)}\right) \\ = \sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{60.3}$$

إذا كان $I(X;Y)$ هو قياس متناظر بين X وY؛ أي X يعبِّر أكثر ما أمكن عن Y، وY يعبر أكثر ما أمكن عن X، فإن :

$$I(X;Y) = H(X)\text{-}H(X|Y) \\ = H(X)+H(Y)\text{-}H(X,Y) \\ = H(Y)\text{-}H(Y|X) \\ = I(Y;X) \tag{61.3}$$

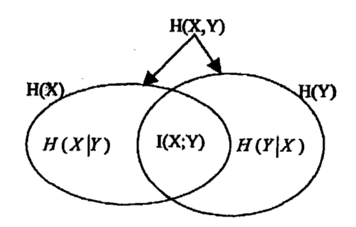
لاحظ أن $I(X;X) = H(X) - H(X|X) = H(X)$، وهكذا المعلومات المتبادلة للمتحول العشوائي X مع نفسه هو أنتروبـــي المتحول العشوائي نفسه، لذلك يسمى الأنتروبـــي

أحياناً بالمعلومات الذاتية. من المثال 2 السابق يمكن بسهولة حساب المعلومات المتبادلة بين المتحولين X وY:

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) = 0.335 \text{ bits}$$

عندما يكون المتحولان مستقلين، فإن $H(X, Y) = H(X) + H(Y)$، ويصبح $I(X; Y) = 0$ في هذه الحالة، ولا يحوي متحول معلومات حول الآخر، وعندما يكون المتحولان X وY مرتبطين فإن $I(X; Y)$ يكون أكبر من الصفر، وعندما يكون الاثنان متماثلين فإن $I(X;Y)$ يكون أعظمياً.

يوضح (الشكل 3.3) مخطط Venn للعلاقة بين H(X) وH(Y) وH (X, Y) وI (X; Y). لاحظ أن المعلومات المتبادلة هي تقاطع المعلومات في X مع المعلومات في Y.



الشكل 3.3: العلاقة بين الأنتروبي والمعلومات المتبادلة

يمكن أيضاً إنشاء العلاقة التالية بين الأنتروبي النسبي والمعلومات المتبادلة بين المتحولين X وY:

$$I(X;Y) = D\big(p(x,y)\|p(x)p(y)\big) \qquad (62.3)$$

مثال 3:

لتكن $\Omega = \{0,1\}$، ولنأخذ التوزيعين p وq على Ω. ليكن $p(1) = r$ و$p(0) = 1 - r$، وليكن $q(1) = s$ و$q(0) = 1 - s$. عندئذ يمكننا حساب:

$$D(p\|q) = (1-r)\log\frac{1-r}{1-s} + r\log\frac{r}{s}$$

وكذلك

$$D(q\|p) = (1-s)\log\frac{1-s}{1-r} + s\log\frac{s}{r}$$

إذا كـــان $r = s$ فإن $D(P\|q) = D(q\|p) = 0$. وإذا كان $r = 1/2$ و $s = 1/4$ نستطيع حساب:

$$D(p\|q) = 1/2\log\frac{1/2}{3/4} + 1/2\log\frac{1/2}{1/4} = 0.2075 \text{ bits}$$

$$D(q\|p) = 3/4\log\frac{3/4}{1/2} + 1/4\log\frac{1/4}{1/2} = 0.1887 \text{ bits}$$

لاحظ أن $D(p\|q) \neq D(q\|p)$ بشكل عام.

الكثير من التعاريف السابقة المعرفة للمتحولات العشوائية المتقطعة تنطبق على الحالة المستمرة لكن يجب أن تؤخذ بحرص عند تعريف الأنتروبـــي للمتحولات العشوائية المستمرة، باعتبار يمكن أن تفرض عدداً غير محسوب من القيم ومن ثم تحمل كمية غير محدودة من المعلومات.

الأنتروبـــي التفاضلي كان مصاحباً للمتحولات العشوائية المستمرة. ليكن X متحولاً عشوائياً بتابع كثافة تجميعي $F(x)$ حيث $F(x) = P(X \leq x)$. إذا كان $F(x)$ مستمراً فإن المتحول العشوائي سيكون مستمراً. أيضاً ليكن $f(x) = F'(x)$ تابع كثافة الاحتمال لــ X (بافتراض أن المشتق معرف)، ولتكن الزمرة، بحيث $f(x) > 0$، هي زمرة دعم لــ X. عندئذ يمكننا تعريف الأنتروبـــي التفاضلي $h(X)$ للمتحول العشوائي المستمر X ذي تابع كثافة الطيف كما يلي:

$$h(X) = -\int_{s} f(x0\log[f(x)]dx \qquad (63.3)$$

حيث S هي زمرة الدعم (Support set) لــ X.

إن تعاريف الأنتروبـــي للمتحولات العشوائية المتعددة المتغيرات المستمرة مشابه للحالة المتقطعة بوضع إشارة التكامل محل إشارة الجمع، وهذا صحيح أيضاً من أجل المعلومات المتبادلة والأنتروبـــي النسبي.

مثلاً، إذا كان $f(x, y)$ كثافة مشتركة للمتحولين العشوائيين X وY، فإن المعلومات

المتبادلة تكون كما يلي:

$$I(X;Y) = \int f(x,y)\log\frac{f(x,y)}{f(x)f(y)}dxdy \qquad (64.3)$$

أيضاً إذا كان f وg هما توابع كثافة للمتحولين العشوائيين المستمرين، فإن الأنتروبـــي النسبـــي $D(f\|g)$ بين الكثافتين f وg يعرف كما يلي:

$$D(f\|g) = \int f\log(f/g) \qquad (65.3)$$

حيث بالتعريف $0\log(0/0) = 0$. من الواضح أنه ليكون $D(f\|g)$ منتهياً يجب أن تكون زمرة الدعم لـــ f محتواة في زمرة الدعم لـــ g.

ومع أن نظرية المعلومات نشأت في ربوع حقل الاتصالات، فقد أُثبت الكثير من العلاقات الهامة بينها وبين حقول أخرى مثل الإحصاء، والآلات الإحصائية، والنظرية العقدية (الحسابات).

سينصب اهتمامنا الرئيسي في نظرية المعلومات لدراسة بنـــى الشبكة العصبونية الصنعية وأدائها عند شروط معينة تُعرَّف بعبارات الأنتروبـــي أو المعلومات المتبادلة على الشبكات. ستعطى أمثلة عن هذه الدراسات في المقاطع الأخيرة من هذا الكتاب.

6.3 مراجعة عامة لنظرية المجموعات العائمة والمنطق العائم

Review of fuzzy set theory and fuzzy logic

اقتُرحت نظرية المجموعة العائمة (الغامضة/المبهمة) من قبل الأستاذ لطفي زاده عام 1965 وذلك في جامعة California-Berkeley، كمحاولة لتعميم نظرية المجموعات التقليدية. ومنذ ذلك التاريخ، توسعت النظرية وامتدت إلى حقول أخرى أسست على نظرية المجموعة، مشتملاً ذلك على المنطق أيضاً.

في المنطق الكلاسيكي (ثنائي القيمة) مثل المنطق التنبئي والافتراضي، تأخذ العبارات أحد تفسيرين أو معنيين اثنين، إما صح أو خطأ. وهذا في الحفظ والصون في نظرية المجموعات التقليدية الواضحة (crisp)، حيث تنتمي الأشياء إلى المجموعة أو لا تنتمي وليس هناك بين بين، أما الوسط فهو مستبعد في نظرية المجموعات التقليدية.

بادئ ذي بدء، المجموعة هي أي مجموعة من الأشياء المعرفة جيداً، ويمكن أن تُجدوَل

(توضع في قائمة list) المجموعة إذا كانت بعدد محدود من الأشياء (العناصر، أعضاء المجموعة). مثلاً، تجدول بمجموعة الأعداد الصغيرة كما يلي: {1, 2, 3, 4}، ويمكن وصف المجموعة بعدد غير منته من العناصر بدون قائمة، مثلاً ({عدد صحيح $x|x$} = Z، {عدد صحيح موجب أو صفر $x|x$ } = N، {عدد حقيقي $x|x$ } = R).

ليكن A وB مجموعتين جزئيتين من المجموعة الكلية (الشاملة) Ω التـــي تحوي n عنصراً. إذا كان x عنصراً ينتمي إلى المجموعة الجزئية A نكتب $x \in A$ وإذا كان x لا ينتمي إلى المجموعة A نكتب $x \notin A$. يمكن أن تعرف المجموعات بخواصها، فإذا كتبنا $\{x|P(x)\}$ نعني العناصر x المحققة لخاصية P.

نقول عن المجموعتين A وB إنهما متساويتان، ونكتب A = B، إذا كان $A \subset B$ و$B \subset A$.

ونقول عن المجموعة A أنها مجموعة جزئية من B ونكتب $A \subset B$ إذا تحقق ما يلي: $(\forall x)(x \in A \Rightarrow x \in B)$، أحياناً يستعمل الرمز $A \subseteq B$، ومن ثم $A \subset B$ يعنـــي أن $A \subseteq B$ و $A \neq B$. وإذا كانت المجموعة A ليست بمجموعة جزئية من B نكتب $A \not\subset B$.

يمكننا الآن استعراض العمليات على المجموعات باختصار، حيث يشير الرمز $\vee$ إلى "أو or" و$\wedge$ إلى "و and" و $\bar{\vee}$ إلى "عدم التكافؤ exclusive or"؛ أي يكون ناتج العملية $P \bar{\vee} Q$ (P,Q متحولات منطقية) صحيحاً إذا كان P أو Q صحيحاً وليس الاثنان معاً، كما يلي:

1. الاجتماع: $A \cup B \equiv \{x \in \Omega; x \in A \vee x \in B\}$
2. التقاطع: $A \cap B \equiv \{x \in \Omega; x \in A \wedge x \in B\}$
3. الفرق: $A \setminus B \equiv \{x \in \Omega; x \in A \wedge x \notin B\}$
4. الفرق التناظري: $A \Delta B \equiv \{x \in \Omega; x \in A \bar{\vee} x \in B\}$
5. المتمم: $A^c, A', CA \equiv \{x \in \Omega; x \notin A\}$

تكون المجموعة A مجموعة مفتوحة إذا كانت A تحتوي فقط على النقاط الداخلية؛ ولا تعتبر النقاط الواقعة على الحدود المحيطة بالمجموعة من المجموعة.

و تكون المجموعة A مغلقة إذا كانت نقاط حدود المجموعة تنتمي إلى المجموعة (هذا يعنـــي أن متمم المجموعة A يكون مفتوحاً).

وتكون المجموعة A محاطة إذا وجد ثابت مثل M يحقق: $|x| < M$ لكل عنصر $x \in A$، حيث يشير | | إلى الطوبلة في الفراغ الإقليدي R^n.

والآن نقول عن الزمرة A أنها متراصة (compact) إذا كانت مغلقة ومحاطة.

يجري تمييز المجموعات من خلال المؤشر (Indicator) أو التابع المميز. يعرف التابع المميز للمجموعة الجزئية A كما يلي:

$$f_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases} \tag{66.3}$$

وبالمثل، يمكن أن يبرهن على أن العلاقات التالية صحيحة لكل قيم x:

$$\begin{aligned} f_{A\cap B}(x) &= f_A(x)\, f_B(x) \\ f_{A\cup B}(x) &= f_A(x) + f_B(x) - f_A(x)\, f_B(x) \\ f_{A\oplus B}(x) &= f_A(x) + f_B(x) - 2f_A(x)\, f_B(x) \end{aligned} \tag{67.3}$$

نترك ذلك للقارئ.

1.6.3 نظرية المجموعة العائمة Fuzzy set Theory

نظرية المجموعة العائمة هي تعميم لنظرية المجموعة الكلاسيكية. إن تعاريف ونظريات وبراهين ونتائج المجموعة الكلاسيكية الواضحة تبقى عموماً صحيحة من أجل نظرية المجموعة العائمة. لقد أُسِّست نظرية المنطق العائم بنفس طريقة المنطق الكلاسيكي في نظرية المجموعة الكلاسيكية ثنائية القيمة. ولكن تحاول محاكمات المنطق العائم محاكاة التمثيل والتعليل المنطقي العقلي للإنسان لمعرفة العالم الحقيقي.

نتعامل في حياتنا اليومية مع الكثير من المفاهيم غير الدقيقة، كالعموميات (مفهوم يطبق على أشياء عديدة)، والالتباسات والمجاهيل، وحوادث الصدفة، والمعرفة غير الكاملة، والمعلومات غير القابلة للتصديق أو حتى المتناقضة. إن تمثيل مثل هذه المفاهيم في نظرية المجموعة التقليدية والمنطق التقليدي صعب جداً إذا لم يكن مستحيلاً.

مثلاً، يصبح الأمر شاقاً بل مزعجاً عندما نحاول في المنطق الكلاسيكي تصنيف الأشياء الموصوفة بتعابير مثل "لا بأس لكن ليس نافعاً" و"أكثر جمالاً بقليل" و"ليس فعلاً طويل" و"أكثر غلاءً بكثير" وهكذا .

تتعامل المجموعة العائمة مع مجموعة جزئية من الكون ليس لها حدود مُعرَّفة تعريفاً كاملاً، وأعضاء المجموعة العائمة سيكون لها درجات تغير الانتماء ومستويات انتساب ضمن المجموعة في المجال [0, 1] (ليس صفراً ولا واحداً كما في الزمرة التقليدية الواضحة).

مثلاً، إن درجة انتماء وائل ذي الطول الفارع 190 سم في المجموعة الجزئية العائمة "الناس الطوال" ستكون قريبة من الواحد، على حين أن درجة انتماء البراء قصير القامة نسبياً 150 سم ستكون أكثر قرباً من الصفر، لكن السؤال الأهم هل عصام ذو الطول 170 سم ينتمي إلى مجموعة الناس الطوال؟ هنا يكمن بيت القصيد. إذا كانت هيفاء فاتنة الجمال فإن درجة انتمائها إلى المجموعة العائمة "النساء الجميلات" ستكون قريبة جداً من الواحد، على حين أن درجة انتماء سارة الفتاة المتواضعة الجمال ستكون أقرب إلى الصفر، ولكننا كيف نستطيع تمثيل درجات انتماء باقي الفتيات إلى هذه المجموعة.

من الضروري استخدام تابع الانتماء، لتشير قيمته إلى درجة ومستوى انتساب كل عضو من أعضاء المجموعة العائمة إلى المجموعة. سنستعمل تعريف التابع المميز للمجموعة العائمة كما فعلنا من أجل المجموعة الواضحة.

لتكن X مجموعة لكون من الأشياء الخاضعة لمفهوم أو اعتبار ما، عندئذ تعرف المجموعة الجزئية العائمة A في X على أنها مجموعة من الأزواج المرتبة:

$$A = \{(x, \mu_A(x))\} \quad , \quad x \in X \qquad (68.3)$$

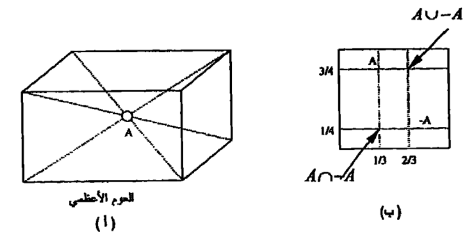
حيث

$$0 \leq \mu_A(x) \leq 1$$

$\mu_A(x)$ هو تابع الانتماء أو التابع المميز الذي يشير إلى درجة انتماء x إلى A. إن قيمة $\mu_A(x) = 0$ تعني أن x لا ينتمي إلى المجموعة A مطلقاً، وقيمة $\mu_A(x) = 1$ تدل على أن x عضو كامل الانتماء في A؛ وهذا موافق لقيمة الانتماء الواضح. إذاً قيم μ بين الصفر والواحد هي الدرجات النسبية لمجال احتواء المجموعة لعناصرها العائمة بين الاحتواء الكامل وعدمه (ومن هنا اشتقت تسمية هذه المجموعة بالعائمة).

يمكننا بالطريقة الهندسية توضيح المجموعة العائمة وذلك بتمثيل هذه المجموعة ككرة نقطية في مركز مكعب واحدي (طول أضلاعه يساوي الواحد) I^n. أحرف المكعب ليست عائمة فهي تمثل القيمة الواضحة ولكن يحصل العوم الأعظمي في النقطة الوسطى للمكعب (أي

نقطة التقاء أقطاره) ذات البعد (0.5 ,0.5 ,0.5) كما هو موضح في (الشكل 4.3أ).



الشكل4.3: التفسير الهندسي للمجموعة العائمة

من هذا الشكل نستطيع تفسير التقاطع والاجتماع للمجموعة العائمة A "ثنائية البعد" والمجموعة المتممة لها A– كما هو موضح في (الشكل 4.3 ب).

مثلاً، إذا عرفنا المجموعة العائمة A بما يلي: A = (1/3, 3/4)؛ أي ينتمي إلى هذه المجموعة عنصران، مثل البراء و وائل، حيث تساوي درجة انتماء البراء إلى المجموعة المبهمة 1/3 (الضلع الأفقي في (الشكل 4.3 ب)) وتساوي درجة انتماء وائل 3/4 (الضلع الشاقولي في (الشكل 4.3 ب) فتكون الزمرة المتممة لها A = (2/3, 1/4)–، ويعطى التقاطع والاجتماع بما يلي:

$$A \cap -A = (\frac{1}{3}, \frac{1}{4}) \quad , \quad A \cup -A = (\frac{2}{3}, \frac{3}{4})$$

العمليات على المجموعة العائمة مشابهة تماماً للعمليات على المجموعة الواضحة المبينة آنفاً، وسنعطي بعض هذه العمليات باختصار، حيث يشير الرمز "–" إلى المتمم:

– تبديلية:

$$A \cup B = B \cup A \qquad (69.3)$$
$$A \cap B = B \cap A$$

– تجميعية:

$$A \cup (B \cup C) = (A \cup B) \cup C \qquad (70.3)$$
$$A \cap (B \cap C) = (A \cap B) \cap C$$

- توزيعية:

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad (71.3)$$

- اللانمو:

$$A \cup A = A$$
$$A \cap A = A \quad (72.3)$$

- الامتصاص:

$$A \cap (A \cup B) = A$$
$$A \cup (A \cap B) = A \quad (73.3)$$

- المتمم والمجموعة الخالية:

$$-(-A) = A \quad , \quad A \cup \Phi = A$$
$$A \cap \Phi = \Phi \quad , \quad A \cap X = A \quad (74.3)$$
$$A \cup X = X$$

- قوانين دو مرغان De Morgan`s laws

$$-(A \cup A) = (-A) \cap (-B)$$
$$-(A \cap B) = (-A) \cup (-B) \quad (75.3)$$

أيضاً نستطيع تعريف العلاقات التالية الخاصة بالمجموعة العائمة:

$$A = B \quad \text{if} \quad \mu_A(x) = \mu_B(x) \quad \forall x \in X$$
$$A \subseteq B \quad \text{if} \quad \mu_A(x) \leq \mu_B(x) \quad \forall x \in X \quad (76.3)$$
$$A \cup B : \mu_{A \cup B}(x) = \max[\mu_A(x), \mu_B(x)]$$

وهكذا $\mu_{A \cup B}$ يمثل المجموعة الجزئية العائمة الصغرى التـــي فيها A وB مجموعات جزئية. وكذلك لدينا:

$$A \cap B : \mu_{A \cap B}(x) = \min[\mu_A(x), \mu_B(x)] \quad (77.3)$$

ويمثل $\mu_{A \cap B}$ المجموعة الجزئية العائمة العظمى التـــي هي مجموعة جزئية لكلا A وB.

أيضاً لدينا:

$$-A : \mu_{-A}(x) \equiv 1 - \mu_A(x) \quad \forall x \in X \quad (78.3)$$

لاحظ عموماً أن:

$$A \cap (-A) \neq \Phi \quad , \quad A \cup (-A) \neq X \qquad (79.3)$$

إذا فرضنا $\mu_A(x) = c$ ، مع العلم أن $0 < c < 1$ ، سيكون لدينا:

$$\begin{aligned} \mu_{A\cup -A}(x) &= \max(c, 1-c) \neq 1 \\ \mu_{A\cap -A}(x) &= \min(c, 1-c) \neq 0 \end{aligned} \qquad (80.3)$$

يمكن إعطاء مثال عن المجموعة العائمة كمجموعة جزئية عائمة A من مجموعة الأعداد الصحيحة الصغيرة. إذا كانت X هي مجموعة كل الأعداد الصحيحة غير السالبة، عندها **يمكننا تعريف** A بواسطة:

$$\mu_A(x) = \frac{1}{1 + (x/4)^2} \quad x = 1, 2, \ldots$$

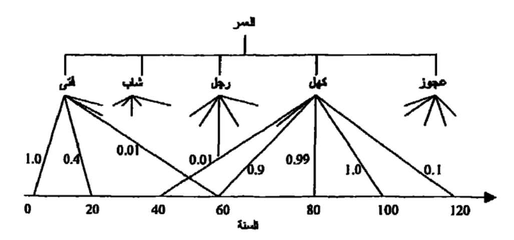
لاحظ في هذا المثال، كان من المقترح أنه "**يمكننا تعريف**..." باعتبار تعاريف تابع الانتماء موضوعية، وهناك دائماً قضية الاختيار الشخصي. بالطبع يستطيع المرء البحث عن إجماع بين مجموعة من الناس في كيفية تعريف توابع الانتماء واستعمال متوسط الإجماع من أجل كل التعاريف.

مثال آخر، ليكن لدينا X مجموعة الأعداد الصحيحة في المجال (0,120) سنة، حيث فسر x بالعمر. بعدئذ يمكننا تعريف المجموعة الجزئية العائمة A "السن" مع قيم الانتماء الموضحة في (الشكل 5.3). لاحظ أن المتحول اللغوي "العمر" يمكن أن يأخذ الكلمات التالية كقيم: فتـــى، شاب، رجل، كهل، عجوز، وكل من هذه المتحولات اللغوية له تابع انتماء عائم كما هو مبين في (الشكل 5.3).

يمكن تعريف بعض العمليات الخاصة التالية على المجموعة العائمة:

- التمدد: Dilation

$$Dil(A) = [\mu_A(x)]^{1/2} \qquad (81.3)$$



الشكل 5.3: تابع الانتماء العائم من أجل "العمر"

– التركيز: Concentration

$$Con(A) = [\mu_A(x)]^2 \quad (82.3)$$

– المعايرة: Normalization

$$Norm(A) = \frac{\mu_A(x)}{\max_x\{\mu_A(x)\}} \quad (83.3)$$

أمثلة أخرى للقياسات العائمة تشمل رئيسي المجموعة (Cardinality) والأنتروبـــي.

يعرف M رئيسي أو "حجم" المجموعة العائمة بما يلي:

$$M(A) = \sum_{i=1}^{n} \mu_A(x_i) \quad (84.3)$$

حيث n عدد عناصر المجموعة.

مثلاً، سيكون رئيسي المجموعة العائمة الممثلة في (الشكل 4.3 ب):

$$M(A) = \sum_{i=1}^{2} \mu_A(x_i) = \frac{1}{3} + \frac{3}{4} = \frac{13}{12}$$

ويعرف أنتروبـــي المجموعة الجزئية العائمة بما يلي:

$$E(A) = \frac{d(A_{near})}{d(A_{far})} \quad (85.3)$$

حيث d المسافة الإقليدية (Euclidean)، وA_{near} وA_{far} هما القطع المستقيمة الواصلة من النقطة A (مثلنا سابقاً المجموعة العائمة A بدائرة نقطية في الفراغ الإقليدي ثنائي البعد) إلى

الوجه المقابل الأقرب والأبعد للمكعب الواحدي I^n على الترتيب. وبالرجوع ثانية إلى (الشكل 5.4 ب) يمكن حساب الأنتروبي E للمجموعة العائمة A كما يلي:

$$E(A)=\frac{d(A_{near})}{d(A_{far})}=\frac{\frac{1}{3}+\frac{1}{4}}{\frac{2}{3}+\frac{3}{4}}=\frac{7}{17}$$

لقد أعطينا هنا أمثلة عن بعض العمليات والتعاريف والنظريات المرتبطة بالمجموعة العائمة. التعاريف الأخرى والعبارات تقع وراء هدف معالجتنا المختصرة هذه. من أجل مزيد من المعلومات يمكن أن يعود القارىء إلى أحد المراجع حول هذا الموضوع.

2.6.3 المنطق العائم Fuzzy Logic

منطق التنبؤ ثنائي القيمة هو المنطق الذي أُسّس على نظرية المجموعة الواضحة التقليدية. تُعرّف التنبؤات صفوف الأشياء أو الأغراض، والأشياء التـــي تحقق تنبؤاً معطى هي أعضاء الصف الموافق.

لقد أُنجزت عدة استنتاجات في المنطق العائم باستعمال قواعد الاستنتاج؛ مثل قاعدة (modus ponens) وقاعدة (modus tollens) وغيرهما. فيما يلي بعض المحاكمات الأساسية الصحيحة التـــي تعتبر قواعد للاستنتاج وتفيد في تبسيط المحاكمات المنطقية وتسهيل عملية التأكد من صحتها أو خطئها، ولكن قبل ذلك علينا إعطاء التعاريف التالية:

المحاكمة المنطقية هي مجموعة منتهية من العبارات المنطقية $p_1, p_2, \ldots, p_n$ التـــي تسمى معطيات أو فرضاً، بالإضافة إلى العبارة q التـــي تسمى نتيجة أو محصلة.

تُمثّل المحاكمة المنطقية كما يلي:

$$\begin{array}{l} p_1 \\ p_2 \\ \vdots \\ \underline{p_n} \\ q \therefore \end{array}$$

أو بالعبارة التالية:

$$p_1 \wedge p_2 \wedge \cdots \wedge p_n \Rightarrow q$$

وتقرأ، إذا كان p_1 صحيحاً و p_2 صحيحاً و...وp_n صحيحاً إذاً ($\therefore$)q صحيح. ويقرأ الرابط الشرطي $\Rightarrow$ بين العبارتين المنطقيتين P وQ ($P \Rightarrow Q$)، إذا كان P فإن Q.

نود التذكير أن قيمة العبارة $P \Rightarrow Q$ خاطئة في حالة واحدة فقط وهي عندما تكون P صحيحة وQ خاطئة. يشـــير الرمز $\vee$ إلى "أو or"، والرمز $\wedge$ إلى"وand"، والرمز $\neg$ إلى المتمم، وسنستخدم الرمز $\rightarrow$ كرابط شرطي في تعبير المعطيات وتعبير النتيجة، وسنستخدم الرمز $\Rightarrow$ كرابط شرطي أيضاً (المكافئ للسطر ــــــــ في تمثيل المحاكمة)، ولكن بين المعطيات والنتائج للتمييز بينهما.

- الجمع: $\dfrac{P}{P \vee Q \therefore}$ أو $\begin{matrix} P \Rightarrow P \vee Q \\ Q \Rightarrow P \vee Q \end{matrix}$

- الاختصار: $\dfrac{P \wedge Q}{P \therefore}$ أو $\begin{matrix} P \wedge Q \Rightarrow Q \\ P \wedge Q \Rightarrow P \end{matrix}$

- اللانمو: $\dfrac{\begin{matrix} P \\ Q \end{matrix}}{P \wedge Q \therefore}$ أو $P \wedge Q \Rightarrow P \wedge Q$

- modus ponens: $\dfrac{\begin{matrix} P \rightarrow Q \\ P \end{matrix}}{Q \therefore}$ أو $P \wedge (P \rightarrow Q) \Rightarrow Q$

- modus tollens: $\dfrac{\begin{matrix} P \rightarrow Q \\ \neg Q \end{matrix}}{\neg P \therefore}$ أو $\neg Q \wedge (P \rightarrow Q) \Rightarrow \neg P$

- disjunctive syllogism:

$\dfrac{\begin{matrix} P \vee q \\ \neg Q \end{matrix}}{P \therefore}$ أو $\neg Q \wedge (P \vee q) \Rightarrow P$

$\dfrac{\begin{matrix} P \vee q \\ \neg P \end{matrix}}{Q \therefore}$ أو $\neg p \wedge (P \vee q) \Rightarrow Q$

- التعدي:
$$\begin{array}{c} P \to Q \\ \underline{Q \to R} \\ P \to R \therefore \end{array} \quad \text{أو} \quad (P \to Q) \wedge (Q \to R) \Rightarrow (P \to R)$$

-
$$\begin{array}{c} P \to Q \\ \underline{R \to S} \\ Q \vee S \therefore \end{array} \quad \text{أو} \quad (P \to Q) \wedge (R \to S) \Rightarrow (Q \vee S)$$

- dilemma
$$(P \vee Q) \wedge (P \to R) \wedge (Q \to R) \Rightarrow R$$
$$\neg P \Rightarrow (P \to Q)$$
$$Q \Rightarrow (P \to Q)$$
$$\neg(P \to Q) \Rightarrow P$$
$$\neg(P \to Q) \Rightarrow \neg Q$$

نقول عن التابع المنطقي إنه كامل (tautology) إذا كانت قيمة التابع "صح" (واحد) مهما كانت قيم متحولاته المنطقية، ويسمى أيضاً بالحقيقة المنطقية، ونقول عن التابع المنطقي إنه تناقض (contradiction) إذا كانت قيمة هذا التابع"خطأ" (صفر) مهما كانت قيمة متحولاته المنطقية. لفهم القواعد المنطقية السابقة والتعاريف سنعطي المثالين التاليين:

مثال 4:

برهن أن قاعدة التعدي هي تابع منطقي كامل (توتولوجي):

$$\begin{array}{c} P \to Q \\ \underline{Q \to R} \\ P \to R \therefore \end{array} \quad \text{أو} \quad (P \to Q) \wedge (Q \to R) \Rightarrow (P \to R)$$

البرهان:

إذا شكلنا جدول الحقيقة للعبارة $P' + Q$ نجد أنه يتطابق مع جدول الحقيقة للعبارة $P \Rightarrow Q$، ولذلك نكتب:

$$(P \Rightarrow Q) = P' + Q \qquad (86.3)$$

بتعويض هذه العلاقة في معطيات ونتيجة القاعدة نحصل على:

$$= (\neg P + Q)(\neg Q + R) \Rightarrow (\neg P + R)$$

وبتعويض العلاقة (86.3) مرة ثانية بين معطيات ونتيجة القاعدة سيكون لدينا:

$$= \neg[(\neg P + Q)(\neg Q + R)] + \neg P + R$$

بفك الأقواس وأخذ المتممات نحصل على:

$$\begin{aligned}
&= \neg(\neg P \neg Q + \neg PR + QR) + \neg P + R \\
&= [\neg(\neg P \neg Q)][\neg(\neg PR)][\neg(QR)] + \neg P + R \\
&= (P + Q)(P + \neg R)(\neg Q + \neg R) + \neg P + R
\end{aligned}$$

وبفك الأقواس والإصلاح سيكون لدينا:

$$\begin{aligned}
&= (P + P\neg R) + QP + Q\neg R)(\neg Q + \neg R) + \neg P + R \\
&= (P + Q\neg R)(\neg Q + \neg R) + \neg P + R \\
&= P\neg Q + P\neg R + Q\neg R + \neg P + R \\
&= \neg P + \neg Q + P + R + Q \\
&= 1
\end{aligned}$$

إذاً القاعدة عبارة عن تابع منطقي كامل. ويمكن إثبات النتيجة نفسها باستخدام جدول الحقيقة، نترك ذلك للقارئ .

مثال 5:

إذا نال عصام الدرجة 80 على الأقل في امتحان مقرر الجبر فإنه سيتخرج هذا العام، وقد نال عصام الدرجة 82 امتحان مقرر الجبر. إذاً سيتخرج عصام هذا العام.

الحل:

لنفترض أن P يمثل العبارة "عصام نال الدرجة 80 في امتحان مقرر الجبر" و Q يمثل "سيتخرج عصام هذا العام". ستكون المعطيات $P \to Q$ و P و Q. فالترتيب المنطقي للمحاكمة سيكون كما يلي:

$$(P \to Q) \wedge P \Rightarrow Q \quad \text{أو} \quad \begin{array}{c} P \to Q \\ \underline{P} \\ Q \therefore \end{array}$$

وهذه المحاكمة هي قاعدة (modus ponens). من جدول الحقيقة التالي نستنتج أن: $P \wedge (P \to Q) \Rightarrow Q$ هي تابع كامل والمحاكمة صحيحة.

P	Q	$P \Rightarrow Q$	$(P \Rightarrow Q) \wedge P$	$(P \Rightarrow Q) \wedge P \Rightarrow Q$
0	0	1	0	1
0	1	1	0	1
1	0	0	0	1
1	1	1	1	1

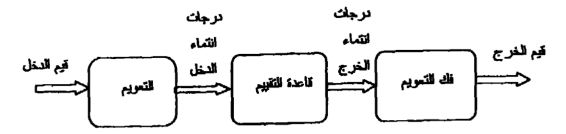
عرِّفت قواعد استنتاج مشابهة للمنطق العائم أُسِّست على علاقات عائمة بين المجموعات الجزئية العائمة، سنهمل التفاصيل هنا وعوضاً عن ذلك سنبين كيف يمكن تفسير القاعدة (modus ponens) العائمة:

- المعطيات: x قيمة صغيرة
- التضمين: x ,y متساويان تقريباً
- النتيجة: y ستكون بقيمة صغيرة أكثر أو أقل.

3.6.3 الأنظمة الخبيرة العائمة Fuzzy expert systems

يَستعمل صنف شائع من الأنظمة الخبيرة قواعد الاستنتاج "إذا كان.. فإن.." لتمثيل كتلة معرفية كبيرة. مثل هذه الأنظمة يعرف بالنظام المبني على قاعدة استنتاج المعرفة (knowledge rule-based system). وقد وظفت الأنظمة الخبيرة الشائعة للعمل بقواعد الاستنتاج "إذا كان.. فإن.." العائمة ودعيت بالأنظمة الخبيرة العائمة. غالباً ما تكون هذه الأنظمة قادرة على تعليل نماذج السلوك تعليلاً أفضل من الأنظمة المبنية على قواعد الاستنتاج الواضحة التقليدية.

يعمل النظام الخبير الأساسي في ثلاث مراحل؛ أولاً مرحلة قلب قيم متحولات الدخل الواضحة إلى قيم مجموعة عائمة (وهذا يدعى بالتعويم)، ومن ثم تطبيق قاعدة التقييم، وأخيراً تحويل قيم المجموعة العائمة عكسياً إلى متحولات الخرج الواضحة (وهذا يدعى بفك التعويم). العملية ككل موضحة في (الشكل 6.3).



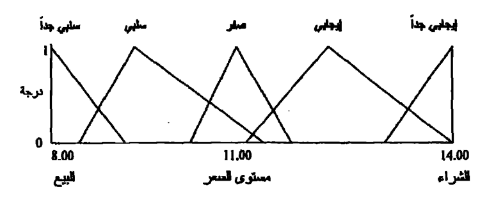
الشكل 6.3: الاستنتاج العائم في نظام خبير بقاعدة عائمة

سنصف طريقة عمل نظام خبير عائم بالمثال التالي: التطبيق العملي الذي اخترناه هو نظام دعم القرار العائم لتجارة الأسهم. لقد عرِّفت توابع الانتماء من أجل مستوى سعر سهم شركة OOB المحدودة. الشركة موضحة في (الشكل 7.3) حيث عرفت قيم الانتماء العائم مثلثية الشكل كما يلي: سلبي جداً، سلبي، صفر، إيجابي، إيجابي جداً.

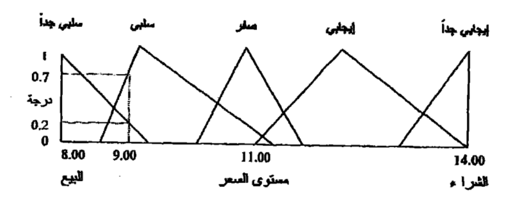
ينجز تعويم مستوى السعر بعملية إسقاط نقاط تقاطع قيمة السعر مع قيم الانتماء العائم على محور تابع الانتماء. مثلاً، إذا كان سعر السهم 9 دولارات فإن درجات الدخل تعطى بالجدول التالي:

قيمة الانتماء العائم	قيمة درجة الانتماء
إيجابي جداً	0
إيجابي	0
صفر	0
سلبي	0.7
سلبي جداً	0.2

وعملية التطبيق (mapping) مبينة في (الشكل 8.3).

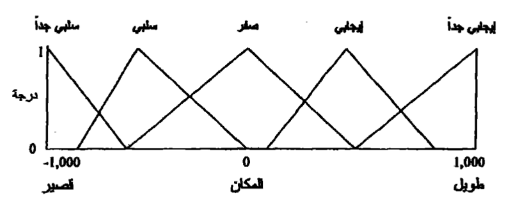


الشكل 7.3: توابع الانتماء لمستوى سعر السهم في شركة OOB المحدودة



الشكل 8.3: عملية تطبيق التعويم

هناك متحول دخل آخر هو مكان السهم، عدد الأسهم المحفوظة (طويل) أو المستحقة (قصير) في قائمة الجرد وفقاً لطموحات شركة التجارة OOB المحدودة. توابع الانتماء لمكان السهم موضحة في (الشكل 9.3).



الشكل 9.3 توابع الانتماء لمكان السهم في الشركة

المرحلة التالية للمعالجة هي قاعدة التقييم. استعملت القواعد العائمة في هذا المثال رابطين شرطيين (الشروط إذا كان.. فإن..) مع متحولين للدخل ومتحول وحيد للخرج. القواعد لها الترتيب التالي:

إذا كان مستوى السعر إيجابياً و مكان السهم إيجابياً جداً

فإن تغير المكان إيجابي.

إذا كان مستوى السعر سلبياً و مكان السهم صفراً

فإن تغير المكان إيجابـــي جداً.

جدول القرار الكامل للقواعد موضحة في الجدول 3-1 حيث السطر والعمود بالكامل هما

شرطا القاعدة، وقيمة الجدول توافق فعل القاعدة.

	مستوى السعر				
	سلبـــي جداً	سلبـــي	صفر	إيجابـــي	إيجابـــي جداً
المكان					
إيجابـــي جداً	إيجابـــي جداً	إيجابـــي جداً	صفر	سلبي	سلبـــي جداً
إيجابي	إيجابـــي جداً	إيجابـــي جداً	إيجابي	سلبي	سلبي
صفر	إيجابـــي جداً	إيجابـــي جداً	إيجابي	صفر	صفر
سلبي	إيجابي	إيجابي	صفر	صفر	إيجابي
سلبـــي جداً	إيجابـــي جداً	إيجابـــي جداً	إيجابي	صفر	صفر

جدول1.3: القرار العائم لدعم قرار تجارة الأسهم

المرحلة الأخيرة من المعالجة هي فك التعويم، أي التطبيق من الجزء "فإن" لتابع انتماء القاعدة إلى قيم متحولات واضحة؛ وهي تمثل الفعل الموظف عليه النظام .

7.3 مراجعة لنظرية الأنظمة غير الخطية والفوضوية

Review of Nonlinear Systems Theory and Chaos

اكتسبت دراسة الأنظمة غير الخطية زخماً معتبراً خلال السنوات العشرين الماضية. ومبعث هذا الاهتمام إدراك حقيقة عدم إمكانية وصف تحولات أو تغيرات ظواهر عديدة بدقة باستعمال النماذج الخطية.

لقد وجدت أمثلة عن الأنظمة غير الخطية في كل مكان، لأنها تحدث في الحوادث الطبيعية والفيزيائية والاجتماعية والمالية وأنواع أخرى من الظواهر المتغيرة التـــي وجدت في الفيزياء عند مستويات الذرة الجزيئية (Gutzwiller عام 1992[14]) وفي الأرصاد الجوية عند المستويات الفيزيائية الكبيرة في الغلاف الجوي.

في الواقع تبدو ظواهر كثيرة للوهلة الأولى أنها عشوائية، ولكنها فعلياً تكون ناشئة عن سلوك نظام غير خطي يدعى بالفوضوي (chaos). وبالفعل فإن السلوكيات غير الخطية تظهر كقاعدة في كثير من الأنظمة المهمة عوضاً عن الاستثناء، ولكن من المؤسف، أن معظم الطرق التحليلية والأدوات الحسابية طُورت لمصلحة الأنظمة الخطية (بالطبع لسهولة دراسة ووصف سلوكيات هذه الأنظمة رياضياً).

في هذا المقطع سنقدم تمهيداً مختصراً لمفاهيم الأنظمة غير الخطية والنظريات ذات الصلة، وذلك لعدة أسباب:

أولاً، إن بنيات الشبكات العصبونية الصنعية الهامة لها تغيرات غير خطية، لا يمكن وصف سلوك هذه الشبكات وصفاً كاملاً دون فهم نظرية الأنظمة المتغيرة غير الخطية.

ثانياً، وجود تطبيقات عديدة للشبكات العصبونية الصنعية مرتبطة بتنبؤ سلوك الأنظمة غير الخطية. مثلاً، لقد أثبتت الشبكات العصبونية الصنعية فعاليتها كأدوات في تنبؤ الحركة المستقبلية للسلسلة الزمنية، التـــي يجب أن تعرف لتقاد بتغيرات غير خطية. وهكذا، يمكن أن يساعد فهم تغيرات السلسلة الزمنية على اختيار أفضل لنماذج الشبكة العصبونية الصنعية.

أخيراً، تبدي الشبكات العصبونية الصنعية نفسها وعوداً تساعدنا على فهم أفضل لتصنيف ووصف الأنظمة غير الخطية، حيث يمكن أن تستعمل استعمالاً فعالاً لتقدير وسطاء النظام ومن ثم تعيننا على وضع مميزات هذه الأنظمة.

يمكن أن توصف تغيرات الأنظمة غير الخطية التـــي نرغب بدراستها بواسطة نظام عام من المعادلات التفاضلية العادية من الشكل:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}) \qquad \mathbf{x} \in \mathbf{R}^n \qquad (87.3)$$

حيث $\mathbf{F}: \mathbf{U} \to \mathbf{R}^n$ و $\mathbf{U}$ مجموعة جزئية مفتوحة من $\mathbf{R}^n$؛ فراغ المتحولات غير المستقلة ويشار له بفراغ الطور أو فراغ الحالة.

تُعرِّف المعادلات حقلاً شعاعياً على جملة مولدة، حيث n قيمة متغيرة ترسم ممراً في فراغ ببعد n كلما تطور النظام مع الزمن. بإجراء تكامل للمعادلة (87.3)، نحصل على حل يعبر عنه بتابع ما، وليكن $\varphi(t)$، من أجل t ضمن مجال زمنـــي ما وليكن I. يدعى الحل $\varphi(t)$ بالمدار (orbit)، والزمرة $\{t\ ,\ \varphi(t) | t \in I\}$ بمسار (trajectory) $\varphi(t)$ في فراغ الطور، ويعرف التطبيق (mapping) (مجموعة كل حلول المعادلة (87.3)) أيضاً بالتدفق (flow).

إذا أخذ الزمن t عند قيم متقطعة فقط، يمكن وصف أو نمذجة تغيرات النظام بمعادلات فروق من الشكل:

$$\mathbf{x}_{n+1} = \mathrm{G}(\mathbf{x}_n) \quad \mathbf{x} \in \mathbf{R}^n \quad n = 0, 1, 2, \ldots \qquad (88.3)$$

حيث $\mathbf{x}_n$ حالة النظام في الزمن n. يدعى الحل φ_n من أجل نظام الزمن المتقطع بالتطبيق

(map). التطبيقات هي تدفقات تشابهية مقطعة عند نقاط في المدار معلّمة بالرقم الصحيح n.

يعتبر مجال اختيار تمثيل النظام في الزمن المستمر أو المتقطع حقلاً واسعاً، حيث يعتمد هذا الاختيار أحياناً على المهمة الموكلة للنظام. عندما يكون زمن الحساب مهماً، يمكن أن يكون نموذج الزمن المتقطع مناسباً أكثر، باعتبار أن الزمن اللازم لتكرار التطبيقات يكون عادة أصغر بمرات عديدة منه في حالة التدفقات. من ناحية أخرى، عندما تكون الدقة هي العامل الأهم، فإن نموذج الزمن المستمر سيكون أكثر مناسبة.

على أية حال، يزداد الحرص في اختيار نظام الزمن المستمر أو المتقطع والقلق من عواقب هذا الاختيار عندما لا تبدي النماذج، عموماً، نفس السلوك (هذا يعنـــي، أن أحد النظامين الزمنيين يمكن أن يتقارب إلى نقطة ثابتة على حين يبقى الآخر مهتزاً).

يمكن أن تصنف الأنظمة التغيرية، مستمرة كانت أو متقطعة، إلى أنظمة محافظة وأنظمة مبددة (conservative /dissipative).

يكون النظام محافظاً إذا حوفظ على الأحجام (الأحجام تبقى ثابتة) في فراغ الطور خلال التطور الزمنـــي للتدفق أو التطبيق. لا يملك مثل هذا النظام مناطق جذب في فراغ الطور، ومن ثم لن تكون هناك نقاط ثابتة ولا حلقات محددة ولا جواذب شاذة (سنشرح الجواذب لاحقاً).

تتميز الأنظمة المبددة، بواسطة تقلص (انكماش) الأحجام في فراغ الطور مع تقدم الزمن. وهذا يعنـــي رياضياً في حالة الأنظمة المعرفة بالمعادلات (87.3) تحقق:

$$\sum_{i}^{n} \frac{\partial F_i(\mathbf{x})}{\partial x_i} < 0$$

للأنظمة المبددة ببعد n. وسيكون فراغ الطور المتقارب مع تقدم الزمن محاطاً بمجموعة جزئية ذات بعد أصغر من n، ويستطيع المرء التمييز بين سلوكيات الدور العابر والدور الطويل أو الدائم للنظام. مثلاً، تكون التدفقات، في نظام مبدد ببعد n مع نقطة توازن مستقرة متقاربة تقارباً كاملاً أو حلقات محددة، متقلصة إما إلى الصفر وإما إلى حالة نهائية أحادية البعد، على الترتيب.

في دراستنا لأنظمة الشبكة العصبونية الصنعية غير الخطية سنركز فقط على الأنظمة

المبددة.

1.7.3 الجواذب Attractors

نحتاج إلى بعض التعاريف لفهم أفضل لسلوكيات النظام المبدد. عموماً، سيكون اهتمامنا فيما يلي منصباً على السلوكيات النوعية للأنظمة غير الخطية المقابلة للسلوكيات الكمية. لذا، لن نبحث عن حلول خاصة أو مدارات للأنظمة، لكن بدلاً من ذلك سنبحث عن معدلات التدفق (تطبيقات) لشروط أولية متغيرة.

تؤدي زمر الجذب والجواذب دوراً هاماً في فهم السلوك المتقارب لهذه الأنظمة، لذا سنبحث في الجواذب أولاً. ليكن φ_t تدفقاً معرفاً على زمرة ما ولتكن U، تكون الزمرة الجزئية $S \subset U$ غير متغيرة إذا كان $\varphi_t(S) \subset S$ لكل قيم $t \geq 0$. وليكن لدينا $A \subset U$ زمرة متراصة (compact)، (الزمرة متراصة compact هي الزمرة المغلقة والمحاطة)، نقول عن الزمرة A إنها زمرة جذب إذا تحقق ما يلي:

1. A غير متغيرة تحت φ_t
2. A لها جوار جذب، هذا يعنـــي وجود جوار مفتوح V من A بحيث $\forall x \in V, \varphi_t(x) \in V$ لكل قيم $t \geq 0$.

أخيراً، نقول عن الزمرة A أنها جاذب (Attractor) للتدفق φ_t إذا كانت زمرة جذب وأيضاً متعدية طبولوجياً (Topological transitivity). يدل التعدي الطبولوجي ضمنياً أنه لأي زمرتين مفتوحتين U وV (حيث $U, V \subset A$) فإن $\varphi_t(U) \cap V \neq \Phi$ لكل قيم t، (هذا يكافئ وجود مدار كثيف على A لـــ φ_t. وهذا يدل ضمنياً أن A لا يمكن أن تجزأ إلى زمرتين غير متغيرتين أو أكثر).

نستطيع التعبير عن الجاذب، تعبيراً غير رسمي، بأنه زمرة من النقاط التـــي تتكدس فيها نقاط المدار المتولد بواسطة التدفق أو التطبيق لقيم كبيرة لـــ t. هذا يشمل حركة متقاربة مستقرة (زمر محددة)، مثل الدوارات والحلقات المحددة المستقرة.

يعرّف حوض الجذب (basin of attraction) بأنه زمرة من كل نقاط الشروط البدائية التـــي لها مدارات تدنو رويداً رويداً لتبقى قرب الجاذب. وبعبارة أخرى، هو مجموعة

المدارات المقبوضة (المأسورة) بواسطة الجاذب. تميِّز الجواذبُ السلوكياتِ النوعيةَ للتدفقات والتطبيقات. ويمكن أن يكون للأنظمة جواذب مستقرة أو غير مستقرة.

يمكن أن تكون الجواذب غير المستقرة في الطبيعة زائدية المقطع أو غير زائدية المقطع (حلزونية الشكل والمقطع).

الجواذب المستقرة أو "الوديعة" هي التـــي من أجلها يحدث التقارب في ممرات مغلقة منتظمة. هناك منطقة مفخخة بالشراك؛ وهي منطقة متصلة اتصالاً بسيطاً ومغلقة في فراغ الطور، بحيث تدخلها كل المدارات خارج هذه المنطقة وعندما تصبح بداخلها لا تخرج منها مطلقاً.

من ناحية أخرى، الجواذب غير المستقرة أو "الشاذة" هي النقاط التـــي يحدث من أجلها سلوك غير منتظم وغير قابل للتنبؤ. تتبع هذه النقاط بممرات متباعدة على نحوٍ واسع أو مسارات غير قابلة للتنبؤ.

بوحه عام، نستطيع تمييز أربعة أنواع أساسية من الجواذب:

1. نقاط ثابتة: وهي نقاط غير متغيرة تحت التطبيق (هذا يعنـــي أنه من أجل النقطة x_n، لدينا $(x_{n+1}=x_n$). والمنحنـــي غير التغيري هو تماماً تعميم للنقطة الثابتة.
2. مدار دوري أو حلقات محددة: وهي جواذب تبدي حركة دورية منتظمة. يعود المسار بدقة إلى نفسه في الزمن T، أي الدور. المدار بدور واحد هو سلوك اهتزازي لتردد واحد، والمدار بدورين اثنين هو سلوك اهتزازي لترددين مختلفين وهكذا دواليك. يمكن أن تكون الحلول الدورية مستقرة أو غير مستقرة.
3. مدارات شبه دورية: السلوكيات شبه الدورية هي حلول تكونت من مجموع حلول دورية بأدوار غير متناسبة (النسبة بين الأدوار ليست جذرية).
4. جواذب فوضوية أو شاذة: إن السلوك المختلط هو حركة متقاربة، أي محاطة. فهي ليست نقطة توازن ولا دورية ولا شـــبه دورية، بل هي حركة لها اعتمادية تتأثر بالشـــروط البدائية SDIC (Sensitive Dependance on Initial Condition) وتتمظهر بسلوك عشوائي. لوحظت الأنظمة الفوضوية بمخططات طيف قدرة واسعة تشبه كثيراً مخططات الأنظمة العشوائية، ومع ذلك فهي تعينية على نحوٍ كامل.

مميزة SDIC تعنـــي أنه من أجل نقطتين أوليتين منفصلتين إحداهما عن الأخرى بمسافة صغيرة جداً، ومع تقدم الزمن ستنفصلان انفصالاً أسيّاً سريعاً جداً حتـــى تصبحا غير مرتبطتين في النهاية. وهذا يجعل من غير الممكن التنبؤ بمثل هذه المسارات، باستثناء حالة الدور القصير. يُستعمل هذا الاعتماد على الشروط البدائية أحياناً كتعريف للجواذب الشاذة أو الفوضوية.

2.7.3 قوى ليابونوف Lyapunov Exponents

من إحدى أفضل الطرائق للتمييز الكمي للأنظمة غير الخطية المتغيرة هي قوى ليابونوف. تعتبر قوى ليابونوف قياساً للمقدار الذي عنده تتباعد/ أو تتعاكس المسارات في كل بعد من فراغ الطور. وبذلك تعطي تعريفاً دقيقاً لـــ SDIC (ومن ثم لفوضوية) النظام .

إذا كانت الحالة البدائية للنظام مضطربة قليلاً (مختلة) في واحدة من الأبعاد n، فإن المقدار الأسي عند هذا الاضطراب يزداد (أو يتناقص) مع الزمن، وهذا ما يسمى بأس أو قوة ليابونوف.

تخيل وجود حجم كروي صغير في فراغ الطور، ستقوم تغيرات النظام المضطرب أو الفوضوي بتشويه الشكل الكروي إلى شكل قطع ناقص بالاتجاهات المتمددة والاتجاهات المتقلصة. يوافق المحور الكبير للقطع الناقص اتجاه عدم الاستقرار الأعظم للتدفق ويقاس مقدار التمدد لهذا المحور بواسطة أس ليابونوف الأعظم. سيتشوه مع الزمن حجم الكرة ليصبح شكلاً معقداً بالكامل. وبتعبير رياضي دقيق، ليكن $r_i(0)$ نصف قطر متناه في الصغر للحجم الأصلي للكرة، و$l_i(t)$ هو طول المحور الأساسي i_{th} للقطع في الزمن t، عندئذ يعرف أس ليابونوف λ_i كما يلي:

$$\lambda_i = \lim_{t\to\infty}\frac{1}{t}\log_2\left(\frac{\ell_i(t)}{r_i(0)}\right) \qquad (89.3)$$

ومن المعتاد أن ترتب قوى ليابونوف تنازلياً $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_k \geq ...$.

تقيس هذه القوى مقدار نمو الفراغات الجزئية في فراغ الطور. وتقيس λ_1 نمو المسافات الخطية السريع مع الزمن.أي نقطتين منفصلتين انفصالاً أولياً بمسافة متناهية في الصغر ε سوف تنفصلان إحداهما عن الأخرى بمقدار $\varepsilon e^{\lambda_1 t}$. وبالمثل، يعين $\lambda_1 + \lambda_2$ المقدار الذي من أجله

تنمو المساحات ثنائية البعد وهكذا.

يكون الجاذب نقطة ثابتة إذا كانت كل قيم λ_i سالبة. وكذلك سيكون لنظام جاذب النقطة الثابتة الثلاثي البعد ثلاث قوى سالبة لليابونوف، وكل الأبعاد الثلاثة تتقلص إلى النقطة الثابتة. سيكون لنظام الحلقة المحددة الثلاثي البعد البسيط قوتان سالبتان وقوة واحدة بقيمة الصفر.

إذا كانت جميع قوى ليابونوف غير سالبة وأصفاراً، يكون الجاذب في هذه الحالة حلقة محددة، وعدد القوى الصفرية يوافق عدد الترددات غير المتناسبة في النظام شبه الدوري.

يكون النظام فوضوياً عندما يكون على الأقل λ_i واحداً موجباً. نموذجياً، ستكون قوى ليابونوف لنظام الجاذب الشاذ الثلاثي البعد على النحو التالي: أس واحد موجب وأس سالب وأس صفري. يميِّز الأس الموجب حساسية النظام المعتمدة على الشروط الأولية وينتج مسارات تتباعد سريعاً (الأس السالب يسبب نقاط متباعدة ولكن تبقى ضمن مجال الجذب).

3.7.3 البعد التجزيئي Fractal Dimension

لتكن لدينا صفيحة ورقية بسماكة مهملة، هذه الصفيحة تعتبر مثالاً للشيء الثنائي البعد. نأخذ هذه الورقة ونجمع بعضها إلى بعض لتشكل كرة ورقية. إنها الآن ليست ثنائية البعد تماماً وليست ثلاثية البعد أيضاً، ولكنها ستكون نوعاً ما بين البعد الثنائي والثلاثي. وسيكون هناك الكثير من الثقوب والفراغات في سطح هذه الكرة. بكلمات أخرى، إنها ببعد جزئي.

التجزيء هو شيء ببعد جزئي، وهو المقابل للشيء بالبعد التكاملي. يعتبر البعد الجزئي أداة هامة أخرى يستطيع المرء بواسطتها إعطاء تمييز كمي للجواذب الفوضوية. نموذجياً، الجواذب الفوضوية لها أبعاد جزئية.

مثال آخر: تخيل حداً متكسراً (مفرَّضاً) في منطقة منبسطة كخط الشاطئ مثلاً. يمكننا طرح السؤال التالي: ما هو بعد هذا الحد؟. إذا كان الشاطئ منحنـياً بنعومة وانسيابية سنميل للقول إنه أحادي البعد. أما إذا كان التكسر في الحد شديداً فسنخصص بعداً أعلى، لكن من المؤكد ليس اثنان. إذا كان البعد ليس واحداً وليس اثنين، فهل نستطيع قياس هذا

البعد؟ لنعتبر التقريب التالي، لنأخذ دوائر بنصف قطر كبير جداً r لتغطي الشاطئ بكامله.

سنقوم الآن بتصغير نصف القطر r تتابعياً قليلاً قليلاً، ومن ثم سنحتاج إلى دوائر بمساحات صغيرة أكثر فأكثر. في النهاية عندما $r \to 0$ فإن عدد الدوائر N(r) سيكون مرتبطاً بالبعد الصحيح D.

لتكن S زمرة النقاط في الفراغ الإقليدي ببعد d، وليكن ε نصف قطر الكرات في هذا الفراغ التـــي نرغب بها تغطية S. إذا كان $N(\varepsilon)$ عدد الكرات اللازمة لتغطية S، عندئذ يعطى البعد D (الجزئي ما أمكن) للزمرة S بالنهاية (مع افتراض وجودها) كما يلي:

$$D \equiv \lim_{\varepsilon \to 0} \frac{\log(N/\varepsilon)}{\log(1/\varepsilon)} \qquad (90.3)$$

إذا كانت S نقطة مفردة فقط فإن كرة نقطية واحدة ستغطيها، ويكون $N(\varepsilon) = 1$ ومن ثم $D = 0$. وإذا كانت S قطعة مستقيمة بطول واحدي، فإن $\varepsilon = 1/N(\varepsilon)$ ومن ثم $D = 1$. وإذا كانت S مستوياً بمساحة واحدية، فإن $\varepsilon^2 = 1/N(\varepsilon)$ ومن ثم $D = 2$ وهكذا.

وهكذا نرى أنه في حالة الأشياء الهندسية المنتظمة، فإن D سيكون لها البعد الإقليدي التكاملي المعتاد. وفي حالة الأشياء الأخرى، فإن D سيكون جزئياً.

المثال المثير للاهتمام هو النظام غير الخطي الثلاثي البعد بجاذب فوضوي جزئي المعطى بالمعادلات التفاضلية الثلاث لـــــ Lorentz عام 1989[15].

$$\begin{aligned}
\frac{dx}{dt} &= \sigma(y(t) - x(t)) \\
\frac{dy}{dt} &= -x(t)z(t) + rx(t) - y(t) \qquad (91.3) \\
\frac{dz}{dt} &= x(t)y(t) - bz(t)
\end{aligned}$$

كان بعد الجاذب يساوي 2.06، وتساوي قيمة قوى ليابونوف من أجل جاذب لورنتز، تقريبياً، $\lambda_1 = 1.37, \lambda_2 = 0.0, \lambda_3 = -22.37$.

هناك قياسات أخرى (أبعاد) مفيدة في تمييز الأنظمة الفوضوية، مثل بعد الارتباط (correlation) وبعد المعلومات (information) وبعد ليابونوف. وهذه كلها قياسات

للفوضى في النظام وهي متعلقة جداً بالبعد الجزئي المشروح آنفاً.

سنهمل التفاصيل هنا ونحيل القارئ المهتم إلى أحد المراجع في موضوع الأنظمة الفوضوية.

4.7.3 بُعْد التضمين والتنبؤ Embedding dimension and Forecasting

لاحظنا في بداية هذا المقطع أن ظواهر عديدة يُعتقد لأول وهلة أنها عشوائية، لكنها كانت تظاهرات فعلية لسلوك فوضوي. باعتبار أن الفوضى تنتج من نظام مقود على نحوٍ تعييني، ومن ثم يمكن أن يكون قابلاً للتنبؤ إذا عرفت طبيعة التغيرات. لكن كيف يمكن كشف تغيرات الأنظمة المعقدة، وخصوصاً تلك الأنظمة ذات درجات الحرية المتعددة؟.

في هذا المقطع الختامي سنتصدى لهذه المسألة، وسنبحث عن التقنيات التـــي يمكن أن تطبق على نحوٍ خاص في النمذجة والتنبؤ بسلسلة الزمن الفوضوية. وسنهتم بالظواهر التـــي تكون قابلة للمراقبة عندما تتطور مع الزمن؛ وقابلة للمراقبة من خلال متحول سلمي وحيد. مثلاً، مؤشرات سعر السوق، مقادير التبادل، مستويات الاحتياطي، غلات صكوك التأمين أو ظواهر مشابهة يمكن أن تكون متجلية على شكل متتالية من المتحولات المفصولة زمنياً بعضها عن بعض بالتساوي $y(t)\,, y(t+\tau)\,, y(t+2\tau)+...$.

نعتقد أن السلسلة الزمنية هي نتيجة نظام متغير مقودٍ بواسطة عدد معين من المتحولات وخاضع لبعض القوانين الديناميكية غير المعروفة. نرغب على الأقل بفهم العلاقات الأساسية، بطريقة نوعية، لأفضل نموذج للنظام لأهداف تنبئيّة. لكن بمتحول واحد فقط (أو بضع متحولات) متوفر للتحليل، كيف يمكننا اكتشاف العلاقات المخفية؟ هذا يبدو مهمة مستحيلة، إذ لا نعرف بَعْدُ النظام وليكن الوحيد من بين المتحولات الهامة. حقاً يمكن أن يكون البعد ضخماً أو حتـــى لانهائياً.

في حالة صنف الأنظمة البسيطة، قد تكون طرق التحليل الخطي كافية لتطوير نموذج مقبول. لقد أثبتت طرق تحليل طيف القدرة لفورييه ومعدل التحرك ذاتـــي الانحدار الإحصائي ARMA (Statistical autoregressive-moving-average) المعروفة جداً بقدرتها على العمل جيداً إذا كانت تغيرات النظام خطية أو تتطور ببطء في فراغ الحالة.

ولكن هذه الطرائق تخفق عندما يكون للمتحولات طيف واسع (كما هو الحال في الأنظمة الفوضوية غير الخطية). ولحسن الحظ تطورت بعض الطرائق غير الخطية لتتعامل مع مسائل أكثر صعوبة، على الأقل عندما يكون لجواذب النظام بُعد منخفض.

إذا انصب اهتمامنا على سلوك الدور الطويل للنظام (ليس السلوك العابر)، فإن معرفة ما بجاذب البعد الأخفض يمكن أن تعطينا كل المعلومات التـــي نحتاج إليها.

إن عملية البحث عن النموذج، ستكون أساساً، إعادة إنشاء فراغ الحالة باستعمال متحول مراقب واحد. توفر النظرية تبعاً لـــ Takens عام 1981[16]، وتطويرها من قبل Mane عام 1981[17] أساساً لطرق إعادة الإنشاء. وهذه النظرية مبنية على براهين هندسية توفر طريقة للذهاب من المتحول المفرد إلى فراغ الحالة المتعدد الأبعاد.

تنص النظرية على أنه من أجل جملة مولدة A متراصة ببعد m، ومن أجل الأزواج (F,g)، حيث F حقل شعاعي و g تابع على A، عندئذ بوجه عام، يكون التطبيق التالي :

$$\Phi_{F,g}(y): A \rightarrow \mathbf{R}^{2m+1} \qquad (92.3)$$

المعرَّف كما يلي:

$$\Phi_{F,g}(y) = [g(y), g(\Phi_1(y),..., g(\Phi_m(y))] \qquad (93.3)$$

هو التضمين (embedding)، حيث Φ_i هو تدفق F.

التضمين هو عملية تطبيق من فراغ أحادي البعد إلى فراغ متعدد الأبعاد، وكذلك اختيار بُعد هذا الفراغ. نرغب بإيجاد التضمين لكي نتعلم أكثر حول تغيرات النظام الأساسي. لاحظ أن التابع g يمكن أن يكون مراقبات متأخرة زمنياً من سلسلة.

تخبرنا النظرية أننا نستطيع إعادة إنشاء سلوك النظام الأصلي بفراغ حالة ببعد m بواسطة تشكيل الشعاع s(n) ذي البعد (2m + 1) من متتالية التأخيرات الزمنية المأخوذة من مراقبات سلسلة أحادية البعد. وهكذا، نستطيع استعمال المراقبات المتأخرة y(n) لتشكيل الشعاع s(n) كما يلي:

$$s(n) = [y(n), y(n+\tau), y(n+2\tau) + ... + y(n+2m\tau)] \qquad (94.3)$$

لتمثيل تغيرات النظام الأساسي.

يجب أن يكون التأخير الزمنـــي τ محدداً، لكن Takens وMane اقترحا أنه يمكن

استعمال أي تأخير زمنـــي. على أية حال، إذا كانت قيمة τ صغيرة جداً فالنقاط المتأخرة لن يتمايز بعضها عن بعض. من ناحية أخرى، إذا كانت قيمة τ كبيرة جـــداً فلن يكون هناك ارتبـــاط (في المعنـــى الإحصائي) بين النقاط. لذا نحتاج إلى بعض الإجراءات لاختيار قيمة متوسطة لـــ τ. والسؤال الآن: كيف نعين قيمة بُعد التضمين d_E؟.

تخبرنا النظرية أنه يجب أن يكون عدداً صحيحاً أكبر بمرتين من بعد الجاذب d_A (الذي لا يلزم أن يكون صحيحاً).

عندما تكون تغيرات النظام معروفة، كما في حالـــة جاذب لورنتز، نختار ببســـاطة البعد $d > 2d_A$؛ فلجاذب لورنتز تكون قيمة $d_E = 5$ باعتبار أن له بعداً جزئياً يساوي 2.06. ويجب أن يقدر d_A عندما لا يكون معروفاً. من الملاحظ هنا أن نظرية التضمين تعطي شرطاً كافياً لـــ d_E، لذا فإن أي بعد d أكبر من $2d_A$ سيكون كافياً.

إن قيمة d_E أكبر من $2d_A$ تضمن أن مدارات المسار ستكون غير منهارة أو متقوضة وستلتف على نفسها في هذه الحالة. ولكن اختيار قيمة d_E بحيث تكون كبيرة جداً سينتج عنه ضجيج غير ضروري وتلوث آخر، ومن ثم سيكون من المستحسن استبعاد هذا الخيار. وأبعد من ذلك، ينمو الحســـاب اللازم نمواً أسياً مع d_E وهذا سبب آخر لإيجاد القيمة الضرورية لـــ d_E.

لقد اقتُرحت طرائق عديدة لإيجاد تقديرات لقيمة d_A. مثلاً، وَصَفَ Abarbanel وزملاؤه عام 1993[18] أربع طرائق لتحليل المعطيات الفوضوية. سنصف إحداها باختصار وهي طريقة الجوار الأقرب الكاذب (**false nearest neighbor**).

استُعملت هذه الطريقة لإيجاد تقدير لـــ d_N البعد الأصغر الضروري لنشر دقيق لتغيرات النظام. إن اختيار $d_E = d_N$ سيعطينا اختياراً أفضل للبعد الذي نعطي به نموذجاً للنظام. قُدِّر البعد d_N باستعمال إجراء يوجِد البعدَ الأصغر الذي يحذف كل التقاطعات الكاذبة للمدارات. تظهر هذه التقاطعات الكاذبة عندما يكون مسقط الجاذب على فراغ صغيراً جداً، وبذلك ستنفصل المسارات المنهارة من قَبل بعضها عن بعض. فمثلاً، إذا أسقطنا جاذب لورنتز الثلاثي البعد على فراغ ثنائي البعد نستطيع رؤية كيف تقوض المدارات بعضها لتنتج تقاطعات كاذبة (أي نقاط تتلامس فيها المسارات أو تكون قريبة جداً بعضها من بعض).

لتحديد الجوارات الكاذبة الناتجة عن الاختيار غير الكافي لحجم البعد d، سنعمل على الشعاع التالي:

$$s(n)=[y(n),y(n+\tau),y(n+2\tau)+...+y(n+(d-1)\tau)] \qquad (95.3)$$

لُبعد معطى d، سيملك الشعاع الجوار الكاذب الأقرب s_{NN}، حيث القرب مبنـــي على المسافة الإقليدية. إذا كانت هذه المسافة صغيرة في حالة البعد d، (ولكن كبيرة نسبياً عندما تحسب في حالة d+1)، نستطيع افتراض مسافة الجوار الأصغر وتكون تبعاً للمسقط من جاذب البعد الأعلى حتـــى البعد الأخفض d. بالذهاب إلى البعد d+1 من d، ستكون النقطتان غير مسقطتين بعيداً إحداهما عن الأخرى، فهما في الحقيقة كانتا في جوار كاذب. لذا فإن d+1 هو البعد الأصغر الذي يسمح بوجود مسارات غير محصورة في فراغ الطور قرب الجاذب.

وسيكون الإجراء بعدئذ ببساطة كما يلي: البداية إيجاد المسافة من أجل s_{NN} في حالة بُعد صغير d، بعدئذ مقارنة المسافة مع المسافة d + 1 وهكذا حتـــى تحدث زيادة حادة في المسافة. عند قيمة عتبة ما، سيحدد هذا القفز في المسافة البعد المقدر d_N. لتكن R_d^2 المسافة التربيعية الإقليدية بين s(n) وs_{NN}، اقترح Abarbanel وزملاؤه عام 1993 طريقة لتقرير متـــى يكون الجوار الأقرب كاذباً فعلاً كما يلي: إذا تجاوزت قيمة النسبة بين المسافة المطلقة بين $s(k+\tau d)$ و $s_{NN}(k+\tau d)$ و$R_d(k)$ قيمة عتبة ما ولتكن R_T مثلاً فإن الجوار الأقرب عند اللحظة الزمنية k سيكون كاذباً، هذا يعنـــي: إذا تحقق

$$\frac{|s(k+\tau d)-s_{NN}(k+\tau d)|}{R_d(k)}>R_T \qquad (96.3)$$

فإن الجوار الأقرب عند اللحظة الزمنية k سيكون كاذباً. عملياً قيم R_T تقع ضمن المجال [10,50]. لكن لسوء الحظ، يعمل هذا المعيار في بعض الحالات فقط، وهذا عيب حقيقي في اقتراح Abarbanel وزملائه.

عند تعيين عتبة المسافة التـــي يحدد بها بعد التضمين، يجب أن نضع في أذهاننا أن القرب هو نسبـــي باعتبار أن مقاييس المسافة نسبية إلى البعد d وحجم الجاذب. لذا لن تكون العتبة مبنية فقط على R_d ولكن أيضاً على حجم الجاذب. إذا قدرنا حجم الجاذب R_A بواسطة

متوسط مربع تباين المراقبات كما يلي:

$$R_A^2 = \frac{1}{N}\sum_{k=1}^{N}(s(k)-\bar{s})^2 \quad , \quad \bar{s} = \frac{1}{N}\sum_{k=1}^{N}s(k) \qquad (97.3)$$

عندئذ فإن المعيار المقترح للجوار الكاذب سيكون معدلاً كما يلي: إذا كان

$$\frac{R_{d+1}(k)}{R_A} \geq 2 \qquad (98.3)$$

عندئذ $s(k)$ وs_{NN} (كلاهما ببعد d) سيصرح عنهما أنهما في جوار كاذب. لذا فإن بعد التضمين الضروري هو $d + 1$. عندما طبقت الطرائق المذكورة آنفاً على جاذب لورنتز، فإن عدد الجوارات الكاذبة انخفض إلى الصفر عند قيمة $d_N = 3$، إضافة لبعض التأكيدات أن هذه التقنية قوية تماماً وعملية.

حتـــى الآن لم نصف طريقة تعيّن زمن التأخير τ. لوحظ مما سبق أنه على الرغم من أن اختيار τ ليس حرجاً في نظرية التضمين، إلا أننا سنعطي معياراً لاختياره. فقد استعملت طرق مختلفة لحساب τ بما في ذلك تابع الارتباط الذاتـــي (autocorrelation)، ولكن ربما يكون أكثر القياسات مناسبة للأنظمة غير الخطية هو الذي أُسّس على استعمال نظرية المعلومات.

يعتبر القياس المبنـــي على المعلومات المتبادلة — بوجه خاص - المكافئَ غير الخطي للقياس الخطي، الارتباط. عندما يكون المتحولان X وY مستقلين بشكل كامل، فإن تابع الاحتمال المشترك لهما $P(x, y)$ يتجزأ إلى تابعي الاحتمال الهامشيين $P(x)\,P(y)$. لذا فإن المعلومات المتبادلة بين هذين المتحولين X وY:

$$I_{XY}(x_i, y_i) = \log_2\left[\frac{P_{XY}(x_i, y_i)}{P_X(x_i)P_Y(y_i)}\right] \qquad (99.3)$$

ستنتهي إلى الصفر كلما أصبح المتحولان أقل ترابطاً.

عندما يكون المتغيران مرتبطين أحدهما بالآخر، عندها تكون المعلومات المتبادلة كبيرة. إن معدل المعلومات المتبادلة بين X وY هو معدل المعلومات المتبادلة المأخوذة عبر كل المراقبات x_i وy_i

$$I_{XY}(n)=\sum_{i=1}^{n}P_{XY}(x_i,y_i)I_{XY}(x_i,y_i) \qquad (100.3)$$

وهذا يقترح استعمال معدل المعلومات المتبادلة كدليل لاختيار τ. في هذا الصدد، إذا أمكننا الحصول على تقديرات توابع الاحتمال المشترك والهامشي لـــ $s(k)$ و$s(k+\tau)$ ولقيم مختلفة لـــ τ، نستطيع حساب معدل المعلومات المتبادلة $I(\tau)$ بين المراقبات.

عند بعض القيم لـــ τ، حيث الارتباط بين $s(k)$ و$s(k+\tau)$ ليس كبيراً ولا صغيراً جداً، يمكن إيجاد زمن التأخير المناسب. بعدئذ سيحدد اختيار التأخير بتقدير الاحتمالات المشتركة والهامشية لـــ $s(k)$ عند الزمن k و$s(k+\tau)$ عند الزمن $k+\tau$ باســـتعمال قيم مختلفة لـــ τ. ويمكن أن يحصل على التقديرات الاحتمالية $P[s(k)], P[s(k+\tau)], P[s(k), s(k+\tau)]$ من المدرجات التكرارية (histograms) الموافقة للنقاط المأخوذة في بعد واحد وبعدين اثنين على الترتيب ومعايرة الحسابات التكرارية.

استَعمل هذا التقريب Abarbanel وزملاؤه لتعيين τ في حالة جاذب لورنتز. فقد اختاروا الأصغر الأول الذي حدث في مخطط معدل المعلومات المتبادلة بين $s(k)$ و$s(k+\tau)$ كتابـــع لـــ t.

سننهي هذا الفصل بالترديد ثانية أن السلوك الفوضوي هو تعيينـــي وقابل للتنبؤ، على الأقل في الدور القصير. يمكن أن توصف الأنظمة الفوضوية باستعمال الطرق المشروحة في هذا المقطع وتميز بالثوابت مثل قوى ليابونوف والأبعاد الجزئية لإنشاء تنبؤ حقيقي للنظام. هذه المفاهيم ستكون مفيدة جداً عندما سنبحث في تطبيقات تنبؤ سلسلة الزمن في الفصول اللاحقة من هذا الكتاب.

الفصل الرابع

أولى الشبكات العصبونية الصنعية وتطورها

عرضنا في الفصل الأول وحدة حساب عصبونية صنعية بسيطة، وأنواع الحسابات التي يمكن أن تنفذها. سندرس في هذا الفصل أقدم بنى الشبكات العصبونية الصنعية وخوارزميات تعليمها، وسنبحث في إمكاناتها وحدود مقدرات مثل هذه الشبكات البسيطة.

أنجزت معظم الأعمال المشروحة في هذا الفصل خلال الأعوام 1950-1960، وهي المدة التي شهدت بحثاً علمياً جدياً، والتي سبقت سنوات الجمود في البحث العلمي التي تلت نشر كتاب Minsly-Papert عن المفسرات ومقدراتها الحسابية المحدودة.

1.4 تمهيد

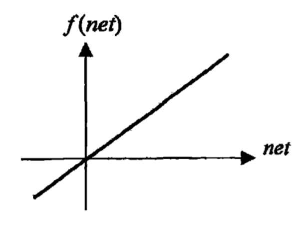
رغم أن معظم الشبكات العصبونية الصنعية التي افترضت ودرست خلال 1950 حتى 1960 كانت شبكات بسيطة بطبقة واحدة، إلا أن النتائج كانت هامة جداً واكتسب منها الباحثون فيما بعد خبرات كثيرة، ووفرت لهم أرضاً خصبةً للبحث والتصميم.

بزغ فجر الشبكات العصبونية الصنعية مع أبحاث McCulloch وPitts عام 1943 على الحسابات العصبونية البيولوجية. وقام Hebb عام 1949 بدراسة الإدراك والتعليم، ومن بعده عمل Rosenblatt عام 1959 على الحواسيب العصبونية الصنعية الأساسية. أما Widrow فقد كان عمله عام 1960 منصباً على العصبونات المتكيفة واستخدامها في التطبيقات الهندسية. وبالطبع لن ننسى الأعمال الهامة للباحثين الآخرين أمثال Steinduch وحتى Kohonen بين أعوام 1963 و1972.

2.4 توابع التفعيل العامة Common Activation Functions

قبل البدء في دراسة الشبكات العصبونية الصنعية وتطبيقاتها لابد لنا من استعراض أهم توابع التفعيل المستخدمة في تلك الشبكات. وكما ذكرنا آنفاً فإن العملية الأساسية للعصبون الصنعي تتحقق بجمع إشارة دخله المثقلة وتطبيق الخرج، أو التفعيل، أو تابع التفعيل.

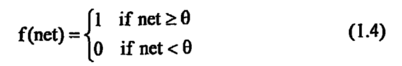
في وحدات طبقة الدخل، يكون تابع التفعيل خطياً $f(net) = net$ ، أي تابع التماثل (Identity) كما هو مبين في (الشكل 1.4). يمكن استخدام تابع التفعيل الخطي نموذجياً في عصبونات أيّ طبقة خاصة في الشبكة العصبونية، مع العلم أن هذا غير ضروري دوماً.



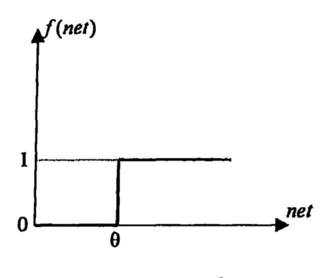
الشكل 1.4: تابع التفعيل الخطي

في معظم الحالات تستعمل توابع التفعيل غير الخطية. لكي نستثمر موارد الشبكات متعددة الطبقات، ولكي تكون مقدراتها العالية مقارنة مع المقدرات المحدودة للشبكات العصبونية الصنعية وحيدة الطبقة، علينا استخدام توابع تفعيل غير خطية (باعتبار نتائج تغذية إشارة خلال طبقتين أو أكثر من عناصر المعالجة الخطية ؛ أي عناصر بتوابع تفعيل خطية،لا تختلف عنها باستعمال طبقة واحدة).

تَستعمل غالباً الشبكاتُ العصبونية الصنعية وحيدة الطبقة تابعَ تفعيل الخطوة الواحدية لعكس دخل الشبكة، الذي يكون بقيمة مستمرة، إلى وحدة الخرج التـــي تكون لها إشارة ثنائية أو ثنائية القطبية. يعرف تابع الخطوة الثنائي أيضاً بتابع العتبة أو تابع Heaviside. يوضح (الشكل 2.4) التابع مع انزياح بمقدار θ .



$$f(net) = \begin{cases} 1 & \text{if } net \geq \theta \\ 0 & \text{if } net < \theta \end{cases} \qquad (1.4)$$



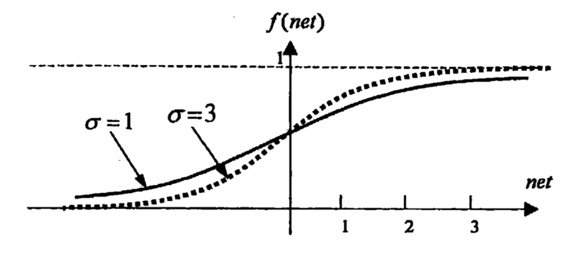
الشكل 2.4 تابع الخطوة الثنائي

تعتبر التوابع Sigmoid (المنحنيات بشكل S) توابع تفعيل مفيدة أيضاً، وأكثر التوابع شهرة واستخداماً هي التابع المنطقي وتوابع الظل القطعي. تستعمل هذه التوابع في الشبكات العصبونية الصنعية المدربة بخوارزمية الانتشار الخلفي، حيث تُخفّف أعباء الحسابات كثيراً خلال التدريب بسبب العلاقة البسيطة بين قيمة التابع عند نقطة وقيمة مشتقه عند نفس النقطة.

يستعمل التابع المنطقي أو تابع Sigmoid ضمن المجال [0,1] غالباً كتابع تفعيل للشبكات العصبونية التـــي تكون لها قيمة الخرج المرغوب به إما ثنائية وإما ضمن المجال [0,1]. للدلالة على مجال التابع يسمى التابع بـــ Sigmoid الثنائي، ويسمى أيضاً بالتابع النسبـــي (logistic) ويعطى بالعلاقة (2-4) مع مشتقه. هذا التابع موضح في (الشكل 3.4) في حالة قيمتين لوسيط الانحدار σ.

$$f(net) = \frac{1}{1 + \exp(-\sigma \cdot net)} \qquad (2.4)$$

$$f'(net) = \sigma \cdot f(net)[1 - f(net)]$$

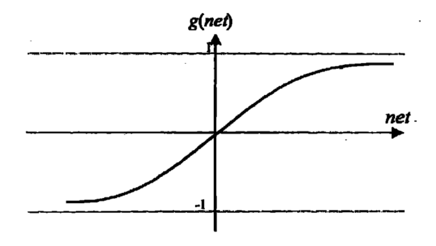


الشكل 3.4: Sigmoid ثنائي

وقد تكون قيمة التابع ضمن المجال [−1,+1]، فيسمى هذا التابع بـــ Sigmoid ثنائي القطبية ، ويعرف بالعلاقة (3.4) مع مشتقه. يوضح (الشكل 4.4) هذا التابع في حالة $\sigma = 1$.

$$\begin{aligned} g(net) &= 2f(net)-1 \\ &= \frac{2}{1+\exp(-\sigma\ net)}-1 \\ &= \frac{1-\exp(-\sigma\ net)}{1+\exp(-\sigma\ net)} \end{aligned} \qquad (3.4)$$

$$g'(\text{net}) = \frac{\sigma}{2}[1+g(\text{net})][1-g(\text{net})]$$



الشكل 4.4 sigmoid ثنائي القطبية

إن التابع Sigmoid ثنائي القطبية قريب جداً من تابع الظل القطعي الذي يعتبر أيضاً من

التوابع الشـــائعة الاستخدام كتابع تفعيل عندما يكون المجال المرغوب به لقيـــم الخرج هو (+1, −1). وسنوضح هذه العلاقة في حالة $\sigma = 1$.

$$g(net) = \frac{1 - \exp(-net)}{1 + \exp(-net)}$$

ويعطى تابع الظل القطعي بالعلاقة:

$$h(net) = \frac{\exp(net) - \exp(-net)}{\exp(net) + \exp(-net)}$$
$$h(net) = \frac{1 - \exp(-2net)}{1 + \exp(-2\,net)} \qquad (4.4)$$

ويعطى مشتق تابع الظل القطعي بالعلاقة:

$$h'(net) = [1 + h(net)][1 - h(net)]$$

من المفضل عادة تحويل المعطيات الثنائية إلى شكل ثنائي القطبية واستعمال تابع Sigmoid أو تابع الظل القطعي. وسنعطي مناقشات أوسع عن اختيار توابع التفعيل وأشكالها المختلفة عند دراسة كل شبكة عصبونية على حدة.

3.4 العصبون التمثيلي أو عصبون McCulloch-Pitts

ربما يكون العصبون التمثيلي هو أول نموذج مقترح لتمثيل عمل العصبون البيولوجي من قبل هذين الباحثين عام 1943[19]. يمكن تلخيص مواصفات هذه العصبونات بما يلي:

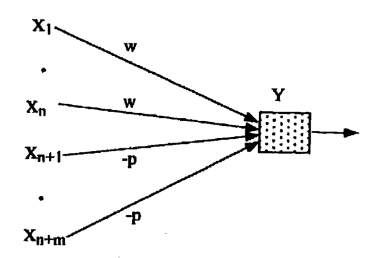
1. إذا كان تفعيل العصبون ثنائياً (0,1)، فهذا يعنـــي أنه عند خطوة زمنية ما سيكون العصبون إما متنشطاً (له تفعيل 1) وإما خاملاً (بتفعيل 0).
2. تتصل العصبونات فيما بينها مباشرة بوصلات مثقلة.
3. تكون الوصلة مهيجة إذا كان الوزن على الوصلة موجباً، وإلا ستكون مخمدة، وكل الوصلات المهيجة إلى عصبون خاص سيكون لها نفس الأوزان.
4. كل عصبون له عتبة ثابتة بحيث إذا كانت قيمة دخل الشبكة net إلى العصبون أكبر من قيمة العتبة عندها سيكون العصبون متنشط.
5. وضعت العتبة بحيث يكون التخميد مطلقاً، وهذا يعنـــي أن أي دخل مخمد ليس صفري

القيمة سيمنع العصبون من التنشط.

6. يستغرق العصبون خطوة زمنية واحدة لكي يمرر الإشارة عبر الوصلة الواحدة.

يوضح (الشكل 5.4) بنية عصبون تمثيلي، حيث يستقبل العصبون Y إشارات الدخل المثقلة بالأوزان. تكون وصلات دخل العصبون مهيجة عندما تكون قيمة الوزن $w > 0$، أو مخمدة عندما يكون الوزن $-p$، حيث $(p > 0)$. سنفرض وجود n دخلاً $x_1, x_2, \ldots, x_n$ يغذي إشارات مهيجة إلى الوحدة Y، وهناك m دخلاً $x_{n+1}, x_{n+2}, \ldots, x_{n+m}$ يغذي إشارات مخمدة. إن تابع التفعيل للوحدة Y هو تابع الخطوة الواحدية:

$$f(net) = \begin{cases} 1 & \text{if} \quad net \geq 0 \\ 0 & \text{if} \quad net < 0 \end{cases} \qquad (5.4)$$



الشكل 5.4: بنية العصبون التمثيلي لـ McCulloch-Pitts

حيث يمثل net إشارة الدخل التركيبية الكلية لعصبون الخرج Y وθ هي العتبة.

يتطلب شرط التخميد أن تكون θ، لتابع التفعيل، محققة للمتراجحة التالية:

$$\theta > nw - mp \qquad (6.4)$$

إن وحدة الخرج Y ستتنشط إذا استقبلت k دخل مهيج أو أكثر، و لم يكن هناك مداخل مخمدة، حيث

$$kw \geq \theta > (k-1)w \qquad (7.4)$$

في البداية، سنعطي بعض الأمثلة لعدد من الشبكات العصبونية التـــي تنجز التوابع المنطقية البسيطة، حيث ســـنفرض أن كل تابع له مدخلان x_1 وx_2، ويعطي قيمة خرج

وحيدة y.

لقد أعطينا قيم الدخل والخرج الشكل الثنائي كتوضيح فقط لما سيتبع، وسنعتمد التحليل عوضاً عن خوارزمية التدريب المستعملة لتعين قيم الأوزان والعتبة. باستعمال هذه العصبونات البسيطة كصناديق بناء نستطيع بناء نموذج لأي تابع أو ظاهرة يمكن أن تمثل بتابع منطقي، لذا سنعطي أمثلة توضيحية نرى من خلالها كيف تُركّب شبكة عصبونية متعددة الطبقات من هذه العصبونات البسيطة لتنفيذ التابع XOR (رياضياً وليس تصنيفاً لأشكال الدخل)، وكيف يُصمَّم نموذج متحسس للحرارة والبرودة الذي سبق أن اقترحه McCulloch-Pitts عام 1943[19].

في هذه الشبكات البسيطة، لكل عصبون عتبة قيمتها 2، ويرمز لتفعيل الوحدة في اللحظة الزمنية t بـــ $x_i(t)$، ويحدد تفعيل العصبون X_i في الزمن t بواسطة التفعيلات في الزمن $t-1$ للعصبونات التـــي يستقبل منها إشارات دخله.

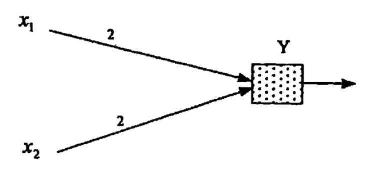
1.3.4 تطبيقات العصبون التمثيلي في حساب التوابع المنطقية

المثال 1:

استخدام العصبون التمثيلي لتنفيذ التابع **AND**، يعطي التابع AND استجابة "صح" إذا كانت قيم المدخلين "صح" وإلا فإن الاستجابة ستكون خطأ. إذا مثلنا الاستجابة صح بالقيمة 1 والاستجابة خطأ بالقيمة 0، وعتبة وحدة الخرج قيمتها 2، فإن أزواج دخل التدريب الأربعة مع الخرج ستكون مبينة في جدول الحقيقة التالي:

الدخل		الخرج
x_1	x_2	y_1
1	1	1
1	0	0
0	1	0
0	0	0

والشبكة التـــي تنجز هذا التابع موضحة في (الشكل 6.4).



الشكل 6.4: العصبون التمثيلي لتنفيذ التابع AND

المثال 2:

استخدام العصبون التمثيلي لتنفيذ التابع **OR**، يعطي التابع OR استجابة صح إذا كانت إحدى قيم الدخل "صح" وإلا فإن الاستجابة ستكون خطأ. هذا (Inclusive OR) باعتبار كلا قيم الدخل يمكن أن تكون "صح" والاستجابة ما تزال "صح". تمثل الاستجابة "صح" بالقيمة 1 والاستجابة خطأ بالقيمة 0، وقيمة عتبة وحدة الخرج تساوي 2، وستكون أزواج دخل التدريب الأربعة مع الخرج مبينة في جدول الحقيقة التالي:

الدخل		الخرج
X_1	x_2	Y_1
1	1	1
1	0	1
0	1	1
0	0	0

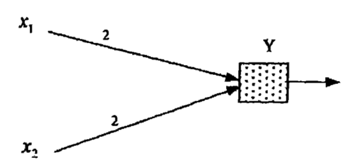
والشبكة المطلوبة موضحة في (الشكل 7.4).

المثال 3:

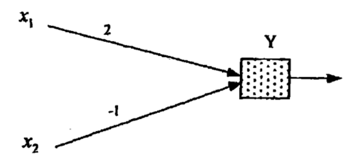
استخدام العصبون التمثيلي لتنفيذ التابع **AND NOT**، التابع AND NOT مثال عن التابع المنطقي غير المتناظر (not symmetric) في معالجته لقيم الدخل. الاستجابة تكون "صح" إذا كانت قيمة الدخل الأول (حصراً) "صح" وقيمة الدخل الآخر (حصراً) خطأ وإلا فإن الاستجابة ستكون خطأ. باستعمال التمثيل الثنائي (0,1) لقيم الدخل والاستجابة، مع أخذ قيمة العتبة لوحدة الخرج قيمتها 2، ستكون أزواج دخل التدريب الأربعة مع الخرج المرغوب به مبينة في جدول الحقيقة التالي:

الدخل		الخرج
x_1	x_2	Y
1	1	0
1	0	1
0	1	0
0	0	0

والشبكة المطلوبة موضحة في (الشكل 8.4). يمكن القول إن العصبون Y يتنشط في اللحظة t إذا وفقط إذا كانت قيمة الدخل $x_1 = 1$ وقيمة الدخل $x_2 = 0$ في اللحظة $t - 1$.



الشكل 7.4: العصبون التمثيلي لتنفيذ التابع **OR**



الشكل 8.4: العصبون التمثيلي لتنفيذ التابع **AND NOT**

مثال 4:

شبكة من العصبونات التمثيلية لتنفيذ التابع **XOR**، التابع XOR (Exclusive or) يعطي استجابة "صح" إذا كان واحد فقط من المداخل "صح" وإلا فإن الاستجابة ستكون خطأ. وباستعمال التمثيل الثنائي وجميع الوحدات Z_1 وZ_2 وY لها عتبة بقيمة 2، ستكون أزواج دخل التدريب الأربعة والخرج كما يلي:

الدخل		الخرج
x_1	x_2	y
1	1	0
1	0	1
0	1	1
0	0	0

يمكن كتابة التابع XOR كما يلي:

$$x_1 \, XOR \, x_2 \leftrightarrow (x_1 \, AND NOT \, x_2) OR (x_2 \, AND NOT \, x_1) \quad (8.4)$$

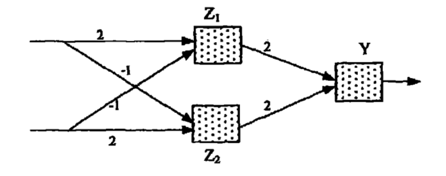
وبالتالي لتنفيذ التابع XOR يلزم أولاً تنفيذ التابعين AND NOT التاليين:

$$\begin{aligned} Z_1 &= x_1 \; AND NOT \, x_2 \\ Z_2 &= x_2 \; AND NOT \, x_1 \end{aligned} \quad (9.4)$$

ومن ثم تنفيذ التابع OR التالي للحصول على التابع XOR:

$$Y = Z_1 \; OR \; Z_2 \quad (10.4)$$

مما سبق نلاحظ أن الشبكة المطلوبة لتنفيذ التابع XOR ستكون مركبة من شبكتين AND NOT يجمع بينهما شبكة OR كما هو موضح في (الشكل 9.4).



الشكل 9.4 شبكة عصبونية لتنفيذ التابع **XOR**

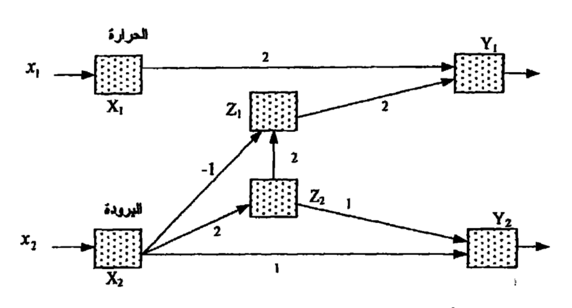
لاحظ أيضاً أن شبكة XOR مؤلفة من طبقتين وليس طبقة واحدة. تتألف الطبقة الأولى من العصبونين Z_1 وZ_2، والطبقة الثانية من العصبون Y.

إن زيادة عدد الطبقات في الشبكة يزيد من قدرتها على تنفيذ مسائل أصعب كحل مسألة XOR غير الخطية في تطبيقات تصنيف أشكال الدخل، كما سنرى لاحقاً.

مثال 5:

نموذج متحسس للحرارة والبرودة باستخدام العصبون التمثيلي لـ McCulloch-Pitts، هناك ظاهرة فيزيائية هامة ومعروفة جيداً هي: إذا طبقت منبهات البرودة على جلد الشخص لمدة زمنية قصيرة جداً، فإن الشخص يشعر بالحرارة. ولكن، إذا طبق على الشخص نفس منبه البرودة لمدة زمنية طويلة، فإن الشخص سيشعر بالبرودة.

باستعمال خطوات زمنية متقطعة تتمكن شبكة من العصبونات التمثيلية لـ McCulloch-Pitts المبينة في (الشكل 10.4) من نمذجة هذه الظاهرة. اقترح هذا النموذج أصلاً McCulloch-Pitts عام 1943[19]. وقد صمم النموذج ليعطي فقط متحسساً واحداً للحرارة أو البرودة المستقبلة بواسطة الوحدات الحساسة.



الشكل 10.4: شبكة McCulloch- Pitts نموذج متحسس للحرارة

مثلت حساسات الحرارة والبرودة في (الشكل 10.4) بالعصبونين X_1 وX_2 على الترتيب، وبالمقابل يمثل العصبونان Y_1 وY_2 أيضاً حساسات الحرارة والبرودة على الترتيب. والعصبونان Z_1 وZ_2 هما وحدتا مساعدة لتنفيذ هذه المهمة.

لكل وحدة في الشبكة قيمة عتبة تساوي 2؛ وهذا يعنـــي أن الوحدة تتنشط (خرجها يصبح 1) إذا كانت قيمة إشارة الدخل أكبر أو تساوي 2.

سيكون دخل الشبكة (1, 0) إذا طبقت الحرارة و(0, 1) إذا طبقت البرودة. والاستجابة

المرغوب فيها للنظام هي الشعور بالبرودة إذا طبقت البرودة على الدخل لمدة خطوتين زمنيتين، هذا يعنـــي أن:

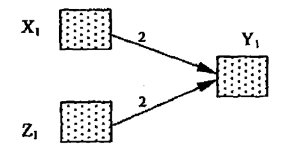
$$y_2(t) = x_2(t-2)\,\text{AND}\,x_2(t-1) \qquad (11.4)$$

إن تفعيل الوحدة Y_2 عند الزمن t هو $y_2(t)$؛ $y_2(t) = 1$ إذا شعر بالبرودة، و $y_2(t) = 0$ إذا لم يشعر بالبرودة.

ومن أجل نمذجة الظاهرة الفيزيائية الموصوفة سابقاً من الضروري أيضاً أن يشعر بالحرارة بتطبيق منبه الحرارة أو منبه البرودة لمدة قصيرة جداً (لمدة خطوة زمنية واحدة). هذا الشرط يمكن التعبير عنه بما يلي:

$$y_1(t) = \{x_1(t-1)\}\,\text{OR}\,\{x_2(t-3)\,\text{AND}\,\text{NOT}\,x_2(t-2)\} \qquad (12.4)$$

لكي نثبت أن الشبكة الموضحة في (الشكل 10.4) هي في الواقع تمثيل للحالتين المنطقيتين السابقتين، سنرى أولاً الوحدات التـــي تعين استجابة Y_1 في الزمن t. وهذا موضح في (الشكل 11.4).



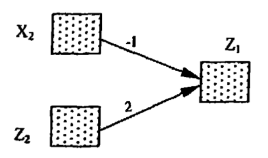
الشكل 11.4: العصبونات التي تعين استجابة الوحدة Y_1

يظهر الشكل أن:

$$y_1(t) = x_1(t-1)\,\text{OR}\,z_1(t-1) \qquad (11.4)$$

سنعتبر الوحدات التـــي تعين استجابة الوحدة Z_1 في الزمن $t - 1$ (وهذا موضح في (الشكل 12.4). حيث من هذا الشكل نرى:

$$z_1(t-1) = z_2(t-2)\,\text{AND}\,\text{NOT}\,x_2(t-2) \qquad (14.4)$$



الشكل 12.4: العصبونات التـــي تعين استجابة الوحدة

أخيراً، استجابة الوحدة Z_2 في الزمن $t - 2$ هي ببساطة قيمة X_2 عند الخطوة الزمنية السابقة:

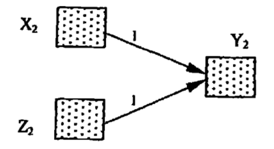
$$z_2(t-2) = x_2(t-3) \quad (15.4)$$

بتعويض التعابير (14.4) (15.4) في (13.4) نحصل على $y_1(t)$ وسيكون:

$$y_1(t) = \{x_1(t-1)\} \text{OR} \{x_2(t-3) \text{AND NOT} x_2(t-2)\}$$

وبالمثل تُجرى للوحدة Y_2 عند الزمن t الإجراءات الســـابقة نفسها. (الشكل 13.4) يظهر أن:

$$y_2(t) = z_2(t-1) \text{AND } x_2(t-1) \quad (16.4)$$



الشكل 13.4: العصبونات التي تعين استجابة الوحدة Y_2

وكذلك سيكون:

$$z_2(t-1) = x_2(t-2) \quad (17.4)$$

وبتعويض (17.4) في (12.4) نحصل على $y_2(t)$ وسيكون:

$$y_2(t) = x_2(t-2) \text{AND} x_2(t-1)$$

وهو المطلوب.

2.3.4 تطبيقات العصبون التمثيلي في تصنيف الأشكال

نجد بسهولة عند الرجوع إلى الفقرة (2.1) من الفصل الأول أن مناطق الاستجابة أو فصل أشكال الدخل للشبكة الممثلة للتوابع المنطقية البسيطة السابقة يمكن أن تخطط كما يلي:

$$y = f(net) \qquad (18.4)$$

حيث f تابع التفعيل، ويكون عادة:

$$f(\text{net}) = \begin{cases} 1 & \text{if} \quad \text{net} \geq 0 \\ -1 & \text{if} \quad \text{net} < 0 \end{cases} \qquad (19.4)$$

و التابع *net* هو تركيب خطي لقيم الدخل المثقلة بالأوزان على الوصلات:

$$\text{net} = w_0 + w_1 x_1 + w_2 x_2 \qquad (20.4)$$

وهذه علاقة مستقيم، ومن ثم سيكون التصنيف بين مناطق صفوف الدخل خطياً.

مثل هذه الشبكات البسيطة لا تستطيع حل المسائل غير الخطية في تصنيف أشكال الدخل (مثل التابع XOR).

لإيجاد خط الفصل بين مناطق الصفوف نضع $net = 0$، حيث يمثل $net \geq 0$ منطقة الصف الأول وسنرمز لها بإشارة "+" و$net < 0$ تمثل منطقة الصف الثانـــي وسنرمز لها بـــ "−".

أحياناً لا يستخدم وزن الانحياز وبدلاً من ذلك تستعمل العتبة بقيمة ثابتة لتابع التفعيل، في هذه الحالة سيكون تابع التفعيل كما يلي:

$$f(\text{net}) = \begin{cases} 1 & \text{if} \quad \text{net} \geq \theta \\ -1 & \text{if} \quad \text{net} < \theta \end{cases} \qquad (21.4)$$

وستكون *net*:

$$\text{net} = w_1 x_1 + w_2 x_2 \qquad (22.4)$$

من العلاقة (20.4) نجد أن معادلة خط الفصل ستكون:

$$x_2 = -\frac{w_1}{w_2} x_1 - \frac{w_0}{w_2} \qquad (23.4)$$

مثال 6:

مناطق الاستجابة للتابع المنطقي **AND**، التابع المنطقي AND من أجل مداخل (x_1, x_2) ثنائية القطبية وخرج y_1 ثنائي القطبية يعرف بجدول الحقيقة التالي:

الدخل		الخرج
x_1	x_2	y_1
1	1	+1
1	−1	−1
−1	1	−1
−1	−1	−1

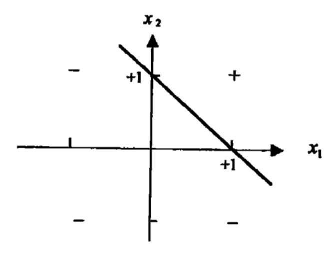
في حالة الأوزان $w_0 = -1, w_1 = 1, w_2 = 1$ ستكون معادلة خط الفصل:

$$x_2 = -x_1 + 1$$

وتُحدَّد إشارة وزن الانحياز w_0 بتحقيق المعادلة:

$$w_0 + w_1x_1 + w_2x_2 < 0 \qquad (24.4)$$

عند قيم الدخل $x_1 = 0 , x_2 = 0$ (يمكن أن تستعمل أي نقطة لا تقع على خط الفصل لتحديد أي طرف من الخط سيكون موجباً وأي طرف سيكون سالباً. نقطة المبدأ مناسبة للاستعمال أيضاً عندما لا تكون واقعة على الخط). يوضح (الشكل 14.4) مناطق الفصل لهذا التابع.



الشكل 14.4: مناطق فصل التابع AND

مثال 7:

مناطق استجابة التابع المنطقي **OR**، إن التابع المنطقي OR من أجل مداخل ومخرج بقيم ثنائي القطبية يعرف بجدول الحقيقة التالي:

الدخل		الخرج
x_1	x_2	Y_1
1	1	+1
1	−1	+1
−1	1	+1
−1	−1	−1

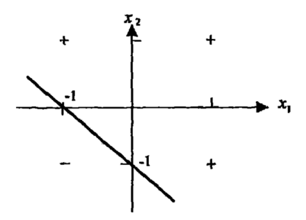
الأوزان يجب أن تكون $w_0 = 1, w_1 = 1, w_2 = 1$ في معادلة الفصل التالية:

$$x_2 = -x_1 - 1$$

واختيار إشارة وزن الانحياز w_0 يتم بتحقيق:

$$w_0 + w_1 x_1 + w_2 x_2 > 0 \qquad (25.4)$$

حيث $x_2 = 0$, $x_1 = 0$، يوضح (الشكل 15.4) إمكانية حد الفصل لهذا التابع.



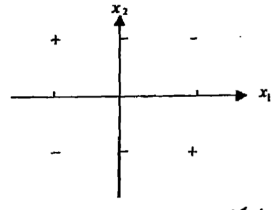
الشكل 15.4: مناطق فصل التابع OR

مثال 8:

مناطق استجابة التابع المنطقي **XOR**، التابع المنطقي XOR لمداخل ومخرج بقيم ثنائية القطبية معرف بجدول الحقيقة التالي:

الدخل		الخرج
x_1	x_2	Y_1
1	1	-1
1	−1	+1
−1	1	+1
−1	−1	−1

ومناطق الاستجابة المرغوب فيها لهذا التابع موضحة في (الشكل 16.4). من هذا الشكل يتضح جلياً أن معادلة خط مستقيم لا تستطيع فصل النقاط التـــي تكون الاستجابة المرغوب فيها (+) ولا (−). لذا سنقوم بحل هذه المسألة فيما بعد بشبكات أكثر تطوراً.



الشكل 16.4 مناطق فصل التابع XOR

4.4 شبكة Hebb

قاعدة تعليم Hebb هي شكل من تعديل الوزن الارتباطي (مرتبط بالقوى الليفية السابقة واللاحقة للعصبون) وتعد هذه القاعدة أقدم وأبسط طرائق تعليم الشبكات العصبونية الصنعية. حيث يُعتبر Hebb أول من اقترح عام 1949[1] أن التعليم يحدث بتعديل القوى الليفية (الأوزان) بطريقة ما، بحيث إذا كان عصبونان متصلان أحدهما مع الآخر داخلياً وكان كلاهما (on) ناشطاً في نفس اللحظة الزمنية t فإن الارتباط بين هذين العصبونين سيزداد قوة.

النص الأصلي للقاعدة يتحدث عن العصبونات المتنشطة عند نفس اللحظة، و لم يذكر أي شيء حول العصبونات المعزَّزة التـــي لا تتنشط عند نفس اللحظة. على أية حال، الوجه الأقوى للتعليم يحدث إذا زدنا أيضاً قيمة الأوزان عندما يكون كلا العصبونين في حالة عدم تنشط " off" في نفس اللحظة. سنناقش ذلك فيما بعد كتوسيع لقاعدة Hebb الأساسية، التـــي وسّعها Rumelharet & McClelland 1986[20].

علل Hebb في النص الأصلي لقاعدته المقترحة عام 1949 أنه عندما يكون محور الخلية A قريباً قرباً كافياً لتهييج الخلية B تكرارياً وبتواصل تبدأ الخلية B بالتنشط، فتحدث عملية نمو ما أو تغير تفاعل حيوي في واحدة أو في كلتا الخليتين، بحيث تزداد فعالية الخلية A عندما

تتنشط واحدة من الخلايا (B).

بالطبع ليس هناك شبكات عصبونية صنعية حقيقية اسمها شبكات Hebb ولكن أطلقت هذه التسمية على الشبكة بطبقة واحدة أمامية التغذية المستعملة لقاعدة تعليم Hebb الموسعة.

إذا كانت المعطيات ممثلة بالشكل الثنائي، فمن السهل التعبير عن تحديث (تعديل الأوزان مع تقدم الزمن) الأوزان بالشكل:

$$w_i^{new} = w_i^{old} + x_i y \qquad (26.4)$$

حيث x_i قيم دخل الشبكة، $i=1,2,...,n$، وy قيمة الخرج. في هذه القاعدة ستكون خوارزمية التعليم كما يلي:

1. إعطاء جميع الأوزان قيماً ابتدائية:

$$w_i = 0 \quad ,i=1,2,...,n$$

2. في حالة كل زوج من شعاع دخل التدريب والخرج المنشود s : t سننفذ الخطوات من 3 إلى 5، حيث s أشعة دخل التدريب $\mathbf{s}=(s_1 ,s_2 ,...,s_n)$.

3. تُفعَّل وحدات الدخل:

$$x_i = s_i \quad ,i=1,2,...,n$$

4. تُفعَّل وحدة الخرج:

$$y = t$$

5. تُعدَّل الأوزان وفقاً لـ:

$$w_i^{new} = w_i^{old} + x_i y \quad ,i=1,2,...,n$$

$$w_0^{new} = w_0^{old} + y$$

تعديل وزن الانحياز ($x_0 = 1$ دائماً)

لاحظ أن وزن الانحياز يعدل تماماً كباقي أوزان الوحدات الأخرى.

يمكن أن يعبر عن تحديث الأوزان بالشكل الشعاعي التالي:

$$\mathbf{W}^{new} = \mathbf{W}^{old} + \mathbf{x}\, y \qquad (27.4)$$

حيث $\mathbf{W}^{new}$ قيمة شعاع الأوزان بعد التحديث، و$\mathbf{W}^{old}$ قيمة شعاع الأوزان قبل التحديث، و$\mathbf{x}$ شعاع الدخل، وy قيمة خرج الشبكة المحسوب.

تكتب هذه العبارة غالباً كما يلي:

$$\Delta W = W^{new} - W^{old} = x\,y \qquad (28.4)$$

من الملاحظ أن هذه الصيغة لا تميز بين زوج التدريب الذي تكون فيه وحدة الدخل on وقيمة الخرج off، وكذلك زوج التدريب الذي تكون فيه وحدة الدخل وقيمة الخرج هما off (لأن الطرف الأيمن للمعادلة (28-4) سيكون مساوياً للصفر). الأمثلة القادمة ستوضح هذه القيود الصارمة لقاعدة Hebb في حالة المعطيات الثنائية، والتنفيذ الجيد سيكون باستعمال معطيات ثنائية القطبية لقيم الدخل والخرج.

1.4.4 تطبيقات شبكة Hebb في تصنيف نماذج الدخل

سنقوم بدراسة بعض التطبيقات المستخدمة لهذه القاعدة، في الصيغ الأصلية لقاعدة تعليم Hebb لم يستعمل مدخل الانحياز على نحو واضح، لذا سيظهر في هذه التطبيقات مدخل الانحياز x_0 كمدخل ثالث له قيمة مساوية دائماً لـــ +1 وبدون هذا المدخل الثالث لا يمكن حل هذه المسائل.

<u>مثال 9:</u>

استخدام شبكة **Hebb** لتمثيل **AND**، بمداخل ومخرج ثنائية، سنمثل المداخل والخرج المرغوب به بقيم ثنائية (1 ,0). إن قيمة الخرج المرغوب به والأوزان وتغير الأوزان (لكل زوج دخل تدريب) تنتج من شعاع الدخل وقيمة الخرج هي كما يلي:

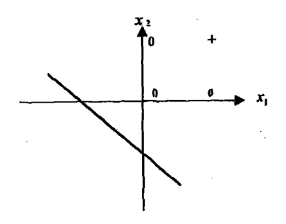
$$w_1^{new} = w_1^{old} + \Delta w_1$$
$$w_2^{new} = w_2^{old} + \Delta w_2$$
$$w_0^{new} = w_0^{old} + \Delta w_0$$

$$\Delta w_1 = x_1 t \quad , \quad \Delta w_2 = x_2 t \quad , \quad \Delta w_0 = x_0 t$$

علينا تدريب الشبكة دوراً واحداً على الأقل لكل أزواج دخل التدريب، وستكون النتائج كما يلي:

الدخل			الخرج	تغير الأوزان			الأوزان		
							w_1	w_2	w_0
x_1	x_2	x_0	T	Δw_1	Δw_2	Δw_0	0	0	0
							قيم ابتدائية		
1	1	1	1	1	1	1	1	1	1
1	0	1	0	0	0	0	1	1	1
0	1	1	0	0	0	0	1	1	1
0	0	1	0	0	0	0	1	1	1

نلاحظ من تحديث الأوزان الموافق لأول دخل زوج تدريب (1, 1, 1) أن خط الفصل سيكون $x_2 = -x_1 - 1$ الموضح في (الشكل 17.4)، حيث يظهر أن الاستجابة ستكون صحيحة فقط في حالة أول نموذج دخل، أما في حالة نماذج التدريب الثانـــي والثالث والرابع



الشكل 17.4: خط الفصل للتابع **AND** الثنائي بعد أول زوج

فلا يحدث تعليم لأن قيمة الخرج تساوي الصفر. وهكذا نجد أن استخدام قيم خرج ثنائية الشكل يمنع الشبكة من تعلم أي نموذج يكون الخرج فيه هو **off** (هذه هي القيود الصارمة للقاعدة في حال استخدام المعطيات الثنائية).

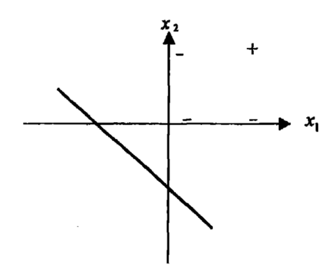
مثال 10:

استخدام شبكة Hebb لتنفيذ **AND**، بمداخل ثنائية ومخرج ثنائي القطبية في هذه الحالة سنحصل على النتائج التالية:

الدخل			الخرج	تغير الأوزان			الأوزان		
							W_1	w_2	w_0
x_1	x_2	x_0	T	Δw_1	Δw_2	Δw_0	0	0	0
							قيم ابتدائية		
1	1	1	1	1	1	1	1	1	1
1	0	1	–1	–1	0	–1	0	1	0
0	1	1	–1	0	–1	–1	0	0	–1
0	0	1	–1	0	0	–1	0	0	–2

ستكون معادلة خط الفصل في حالة تحديث الأوزان الموافق لأول زوج تدريب:

$x_2 = -x_1 - 1$، وهي موضحة في (الشكل 18.4)، حيث يُظهر هذا الشكل أن استجابة الشبكة صحيحة فقط في حالة أول نموذج دخل، والتعليم يستمر في حالة النموذج الثاني والثالث والرابع باعتبار قيمة الخرج هي –1 بدلاً من الصفر في المعطيات الثنائية.



الشكل 18.4: خط الفصل للتابع **AND** الثنائي بعد أول زوج تدريب

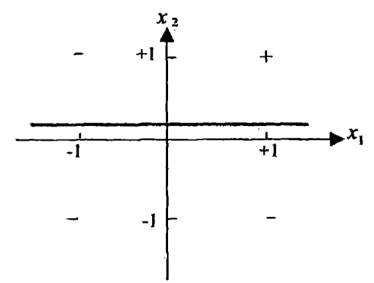
على أية حال، يؤدي اختيار نماذج التدريب دوراً هاماً في تحديد المسائل التي يمكن حلها باستعمال قاعدة Hebb.

مثال 11:

استخدام شبكة **Hebb** لتنفيذ التابع **AND**، بمداخل ومخرج ثنائية القطبية في هذه الحالة سنحصل على النتائج التالية:

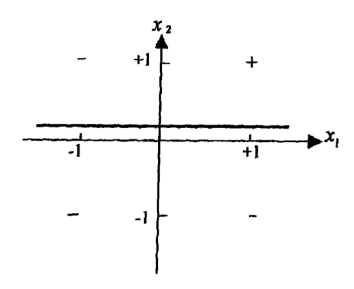
x_1	x_2	x_0	T	Δw_1	Δw_2	Δw_0	W_1	w_2	w_0
الدخل			الخرج	تغير الأوزان			الأوزان		
x_1	x_2	x_0	T	Δw_1	Δw_2	Δw_0	0	0	0
							قيم ابتدائية		
1	1	1	1	1	1	1	1	1	1
1	−1	1	−1	−1	1	−1	0	2	0
−1	1	1	−1	1	−1	−1	1	1	−1
−1	−1	1	−1	1	1	−1	2	2	−2

يمثل (الشكل 19.4) معادلة الفصل $x_2 = -x_1 - 1$ لأول شعاع دخل تدريب.



الشكل 19.4: خط الفصل للتابع **AND** الثنائي بعد أول زوج تدريب

يظهر من الشكل أن استجابة الشبكة ستكون صحيحة في حالة نموذج الدخل الأول فقط. وفي حالة زوج التدريب الثاني فإن معادلة خط الفصل ستكون $x_2 = 0$ الموضحة في (الشكل 20.4). يظهر من هذا الشكل أن استجابة الشبكة ستكون صحيحة في أول نموذجين للدخل، وعَرَضياً للنموذج (1-, 1-). أما أزواج التدريب الثالث والرابع فإن استجابة الشبكة ستكون صحيحة تماماً، كما هو مبين في (الشكل 21.4) لمعادلة الفصل لكلا نموذجي الدخل.

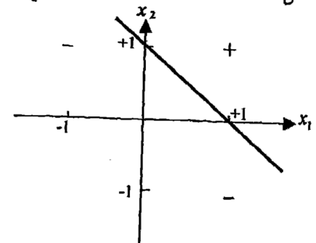


الشكل 20.4: خط الفصل للتابع AND الثنائي بعد زوج التدريب الثاني

2.4.4 تطبيقات شبكة Hebb في تعرف الأشكال

مثال 12:

شبكة **Hebb** لتصنيف أشكال الدخل ثنائية البعد (تمثيل الأحرف) يمكن استخدام شبكة Hebb أو قاعدته في تعرّف الأشكال عموماً، وفي تعرّف الأحرف(Character recognition) خصوصاً، لذا سنستعمل هذه القاعدة في تعرّف حرف واحد، ويتم ذلك بتدريب الشبكة لتعرّف الشـــكل "X" والشكل "O". مُثّلت هذه الأشـــكال كما يلي:



الشكل 21.4: خط الفصل للتابع AND الثنائي بعد زوج التدريب الثالث

النموذج الثاني " O "	النموذج الأول" X "
0 # # # 0	# 0 0 0 #
# 0 0 0 #	0 # 0 # 0
# 0 0 0 #	0 0 # 0 0
# 0 0 0 #	0 # 0 # 0
0 # # # 0	# 0 0 0 #

حيث الرمز "#" يمثل نقطة مضاءة من الحرف، والرمز "0" يمثل نقطة غير مضاءة من الحرف. لمعالجة هذا المثال كمسألة تصنيف شكل مع خرج وحيد للشبكة (صف واحد للخرج)، سنعتبر أن الصف المنشود هو "X" وسنأخذ الشكل "O" ليكون الشكل الذي ليس هو الشكل المنشود " X".

علينا أولاً تحويل الأشكال إلى أشعة دخل بقيم ثنائية القطبية، يتم ذلك بسهولة بتخصيص كل رمز "#" بقيمة 1+ وكل رمز "0" بقيمة 1− للحصول على معطيات ثنائية القطبية. تتم عملية التحويل هذه للشكل ثنائي البعد (الحرف) إلى شعاع دخل سطراً تلو سطر؛ هذا يعنـــي أن المكافئ للسطر الثانـــي من الشكل يأتـــي بعد مكافئ السطر الأول، والثالث بعد الثاني، وهكذا.

يصبح الشكل الأول بعد التحويل ممثلاً بشعاع السطر التالي:

(1 −1 −1 −1 1, −1 1 −1 1 −1, −1 −1 1 −1 −1, −1 1 −1 1 −1, 1 −1 −1 −1 1)

ويصبح الشكل الثانـــي بعد التحويل ممثلاً بشعاع السطر التالي:

(−1 1 1 1 −1, 1 −1−1 −1 1, 1 −1 −1 −1 1, 1 −1 −1 −1 1, −1 1 1 1− 1)

لاحظ أنه سيكون للشبكة $5 \times 5 = 25$ وحدة دخل ومخرج وحيد.

عند استخدام الحاسوب يمكن كتابة البرنامج بحيث تجري قراءة الشعاعين باستعمال التوصيف (Format) الثنائي البعد.

ستكون الاستجابة الصحيحة في حالة الشكل الأول هي "on" أو 1+، وهكذا بعد تمثيل النموذج الأول فإن الأوزان ببساطة هي النموذج الأول نفسه x ، ووزن الانحياز بعد هذا التمثيل هو 1+.

والاستجابة الصحيحة للشكل الثانـــي هي " off" أو 1−، وهكذا فإن تغير الوزن عندما

يقدم النموذج الثانـــي للشبكة سيكون:

(1 −1 −1 −1 1, −1 1 1 1 −1, −1 1 1 1 −1, −1 1 1 1 −1, 1 −1 −1 −1 1)

إضافة إلى ذلك سيكون تغير وزن الانحياز هو 1−، وبإضافة تغير الأوزان إلى الأوزان المثلة للنموذج الأول نحصل على الأوزان النهائية:

(2 −2 −2 −2 2, −2 2 0 2 −2 ,−2 0 2 0 −2, −2 2 0 2 −2, 2 −2 −2 −2 2)

ويصبح وزن الانحياز 0. نحسب الآن خرج الشبكة لكل نماذج التدريب. دخل الشبكة لأي نموذج تدريب هو جداء سلمي بين نموذج الدخل مع شعاع الوزن $. net = \mathbf{x}^T \cdot \mathbf{w} = \sum_{i=1}^{25} x_i w_i$

ففي شعاع التدريب الأول، سيكون دخل الشبكة مساوياً لـ 42 وهكذا نجد أن الاستجابة موجبة وهو المطلوب. وفي نموذج التدريب الثاني، دخل الشبكة يساوي 42− وتكون الاستجابة سالبة وهذا هو المطلوب أيضاً. إذاً قامت الشبكة بتعرّف الأحرف، حيث استطاعت تمييز أي حرف قدم لها من خلال استجابتها الموجبة أو السالبة.

على أية حال، تستطيع الشبكة إعطاء استجابات معقولة لنماذج دخل مشابهة لنماذج التدريب ولكن ليست متماثلة (نماذج دخل ضجيجية). هناك نوعان من التغيرات يمكن إحداثها في أحد نماذج التدريب لتوليد نماذج دخل جديدة ضجيجية، ويتوقع أن تكون استجابة الشبكة صحيحة.

النوع الأول للتغير يشار له عادة بعبارة "أخطاء في المعطيات"، في هذه الحالة يجري تغير إشارة مُركّبة أو أكثر من شعاع الدخل (الموافقة لنقطة أو أكثر (pixel) في النموذج الأصلي) من 1+ إلى 1− أو بالعكس. النوع الثانـــي للتغير ويدعى "معطيات ضائعة"، في هذه الحالة ستأخذ مركبة واحدة أو أكثر من شعاع الدخل قيمة الصفر بدلاً من 1+ أو 1− (سنقوم بإعطاء بعض الأمثلة عما سبق لاحقاً).

<u>**مثال 13:**</u>

قيود تدريب قاعدة **Hebb** للنماذج الثنائية، في هذا المثال سيظهر أن قاعدة Hebb يمكن

أن تخفق حتـــى إذا كانت المسألة عملية فصل خطي، وحتـــى إذا لم يكن الخرج صفراً. سنعتبر الدخل والخرج كالآتـــي:

الدخل			الخرج
x_1	x_2	x_3	t
1	1	1	1
1	1	0	0
1	0	1	0
0	1	1	0

من السهل إثبات أن قاعدة Hebb لا تستطيع تعلم أي نموذج يكون الخرج فيه مساوياً الصفر. لذا يجب علينا على الأقل عكس قيم الخرج إلى قيم ثنائية القطبية (+1, −1)، كما هو معطى في الجدول التالي:

الدخل			الخرج
x_1	x_2	x_3	t
	1	1	1
1	1	0	−1
1	0	1	−1
0	1	1	−1

يبين (الشكل 22.4) أن المسألة أصبحت الآن قابلة للحل، وهذا يعنـــي أن نقاط الدخل المصنفة في صف واحد (من أجل قيمة الخرج +1) ستكون مفصولة خطياً عن تلك التـــي لا تنتمي إلى الصف (بقيمة خرج −1).

يظهر الشكل ضرورة وجود الانحياز غير الصفري، باعتبار أن مستوى الفصل لا يمر من نقطة المبدأ. تعطى معادلة مستوى الفصل كما يلي:

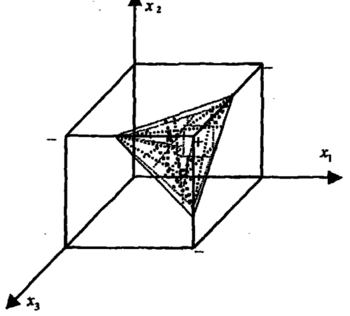
$$x_1 + x_2 + x_3 + (-2.5) = 0$$

وهذا يعنـــي أن شعاع الوزن قيمته (1, 1, 1) وقيمة وزن الانحياز هي −2.5.

شُكِّلت الأوزان مع الانحياز بأخذ مجموع تغيرات الأوزان التـــي تحدث عند كل مرحلة من الخوارزمية. إن تغير الوزن هو ببساطة نموذج الدخل مضروباً بقيمة الخرج لهذا النموذج (ستكون لدينا مركبة رابعة تمثل دخل الانحياز، والذي يأخذ القيمة واحد دوماً).

وبالتالي سنحصل على النتائج التالية:

تغير الوزن لنموذج الدخل الأول	1	1	1	1
تغير الوزن لنموذج الدخل الثانـــي	−1	1	0	−1
تغير الوزن لنموذج الدخل الثالث	−1	0	−1	−1
تغير الوزن لنموذج الدخل الرابع	0	−1	−1	−1
الأوزان النهائية والانحياز	−1	−1	−1	−2

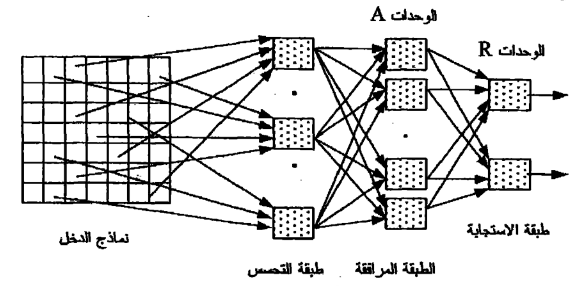


الشكل 22.4: مستوى الفصل لنماذج التدريب الثنائية

5.4 شبكات perceptrons البسيطة

كان عالم النفس Frank Rosenblatt أول من اقترح شبكة المفسر (perceptron) وذلك في منتصف 1950[21]. قصد من هذه الشبكات أن تكون نماذج حسابية لشبكية العين. كان دخل الشبكة عبارة عن مصفوفة من الحساسات الضوئية المرتبة بشكل شبكية مربعة.

كانت الحساسات متصلة عشوائياً (بالمعنـــى المكاني) بطبقة الوحدات المرافقة A (Association units)، من خلال طبقة التحسس التـــي تحول نماذج الدخل إلى الشكل الثنائي، والتـــي بدورها تتصل بطبقة وحدات الاستجابة R (Response units) اتصالاً كاملاً. كان الهدف من النظام تفعيل وحدة الاســـتجابة لنماذج دخل معطاة. يوضح (الشكل 23.4) بنية شبكة perceptron بسيط.



الشكل 23.4: شبكة perceptron بسيطة

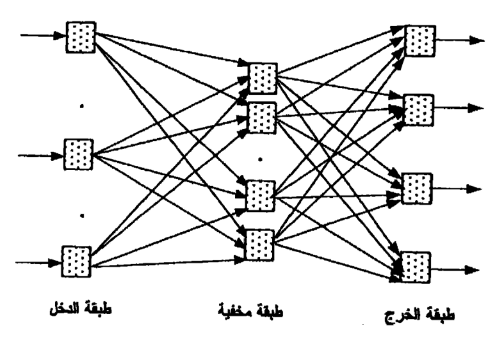
تُستعمل الوحدات A تابع التفعيل الخطي $f(net) = net$ ، ويرسل خرج هذه الوحدات إلى الوحدات R من خلال وصلات مثقلة بأوزان قابلة للتعديل. يساوي خرج الوحدة R الصفر إذا كان مجموع المداخل المثقلة أصغر أو يساوي الصفر، ويساوي المجموع المثقل (لأنه تابع خطي) إذا كانت قيمة الدخل أكبر من الصفر. في نماذج أخرى للشبكة، الخرج هو تابع القفزة الواحدية بقيم ثنائية (0, 1) أو ثنائية القطبية (1–, 1+).

خوارزمية تدريب الشبكة هي نوع من التعليم بمعلم، بحيث تعدل الأوزان لإنقاص قيمة الخطأ في خرج الشبكة عندما لا يكون منسجماً مع الخرج المنشود.

تعتبر قاعدة تعليم البيرسبترون أقوى من قاعدة تعليم Hebb، وقد أثبت Rosenblatt أن هذه الخوارزمية متقاربة دائماً (حيث يُكرَّر التعليم حتـــى توجد زمرة من الأوزان التـــي تسمح للشبكة بإعطاء قيمة صحيحة للخرج لجميع نماذج التدريب على الدخل) تحت شروط معينة.

ودرس Rosenldalt أيضاً صنفاً من الشبكات ذات التعليم بدون معلم أو ذات التكيف الذاتـــي (Self-adaptive). يوضح (الشكل 24.4) أنموذجاً لشبكة بيرسبترون عامة بطبقات متعددة، حيث أجريت تعديلات عديدة على الشبكة الأساسية لـــ Rosenblatt ودرست عبر السنين التالية من اقتراحها.

تطبق هذه الأنظمة الآن عملياً كوحدات لتعرّف الأشكال، حيث دربت بخوارزميات جديدة معدلة عن الخوارزمية الأساسية للبيرسبترون من قبل Gallant عام 1993[22] أسماها خوارزمية المحفظة (pocket) و(rachet-pocket) التـــي سنشرحها لاحقاً.



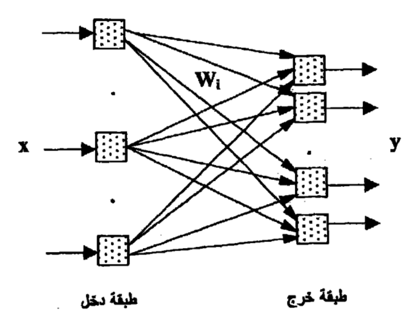
الشكل 24.4: شبكة بيرسبترون متعدد الطبقات بتغذية

1.5.4 الخوارزميات الأساسية لتعليم البيرسبترون
Basic perceptron learning algorithms

تتألف شبكة بيرسبترون بسيطة من طبقة واحدة بتغذية أمامية ووحدات منطقية مع عتبة. يوضح (الشكل 25.4) البنية المكافئة للبيرسبترون، حيث أعطيت نماذج التدريب مع المداخل والخرج قيماً ثنائية (0, 1) وطبقة الوصلات الوحيدة بين طبقة الدخل والخرج كانت أوزانها بقيم حقيقية قابلة للتعديل، وقد مُثّلت هذه الأوزان بالمصفوفة **W**.

على الرغم من وجود نسخ مختلفة لشبكة البيرسبترون وقواعد تعليم موافقة لها، فإن

القاعدة الأساسية للتعليم هي تبديل الأوزان فقط عندما يوجد خطأ بين خرج الشبكة المحسوب y والخرج المنشود أو المثالي t. في هذه الحالة تزداد قيمة الأوزان على الخطوط الفعالة (المداخل مساوية للواحد) بمقدار صغير إذا كانت قيمتها يجب أن تصبح مساوية للواحد (فعالة) وكانت قيمتها الفعلية تساوي الصفر (غير فعالة)، وتخفض قيمة هذه الأوزان بمقدار صغير عندما تكون الخطوط مساوية للواحد (فعالة) ولكن يجب أن تصبح بقيمة الصفر (غير فعالة). يمكن أن يكون مقدار الزيادة والنقصان قيمةً ثابتةً أو قيمةً متناسبةً مع جداء الخطأ وتفعيل الدخل.



الشكل 25.4: شبكة بيرسبترون بسيطة

يمكن تلخيص قاعدة تعليم البيرسبترون كما يلي:

1. إذا كان الخرج فعالاً (1) وسيبقى فعالاً أو كان غير فعالٍ (0) وسيبقى غير فعال، في هذه الحالة لا يحدث تعليم (لا تغير في قيم الأوزان).
2. إذا كان الخرج غير فعالٍ (0) وسيصبح فعالاً (1)، عندها تزداد قيم الأوزان على كل خطوط الدخل الفعالة.
3. إذا كان الخرج فعالاً (1) وسيصبح غير فعال (0)، عندها تنقص قيم الأوزان على كل خطوط الدخل الفعالة.

يمكن التعبير عن هذه الخوارزمية بالعلاقة الرياضية التالية:

$$\mathbf{W}^{new} = \mathbf{W}^{old} + \Delta\mathbf{W} \qquad (29.4)$$

حيث تمثل المصفوفة $\Delta\mathbf{W}$ تغيّر قيم عناصر مصفوفة الأوزان، ويعبر عن عناصرها كما يلي:

$$\begin{aligned} &\Delta w_{ij} = \alpha \left(t_j^p - y_j^p\right) x_i^p = \alpha \times error \times input \\ &i = 1,2,\ldots,n \quad ,j = 1,2,\ldots,m \end{aligned} \qquad (30.4)$$

في المعادلة (30.4) Δw_{ij} يمثل تغير الأوزان على الوصلة بين خرج وحدة الدخل رقم i ودخل وحدة الخرج رقم j، و x_i^p يمثل مركبة الدخل رقم i، حيث $i = 1,2,\ldots,n$، الموافقة لنموذج التدريب رقم p للوحدة، و t_j^p قيمة الخرج المنشود للوحدة رقم لمركبة الدخل x_i^p، و y_j^p هو الخرج المحسوب بالعصبون رقم j عندما يكون الدخل x_i^p، و$\alpha > 0$ وسيط معدل التعليم.

وهكذا، في حالة دخل التدريب رقم $k + 1$، ستكون قيمة الوزن w_{ij} حيث $i = 1,2,\ldots,n$ و $j = 1,2,\ldots,m$ معدلة (محدثة) كما يلي:

إذا كان الخرج 0 ويجب أن يكون 1: $\quad w_{ij}(k+1) = w_{ij}(k) + \alpha\, x_i$

إذا كان الخرج 1 ويجب أن يكون 0: $\quad w_{ij}(k+1) = w_{ij}(k) - \alpha\, x_i \qquad (31.4)$

إذا كان الخرج صحيحاً (لا تغير): $\quad w_{ij}(k+1) = w_{ij}(k)$

لاحظ أن الأوزان تعدل قيمتها بالزيادة أو النقصان لكي يُزاح دخل الشبكة *net* باتجاه قيم الأوزان اللازمة لإعطاء خرج صحيح.

يحدد معدل التعديل بواسطة العامل α، إذا كان معدل التعليم α صغيراً جداً فإن التعليم سيجري ببطء ولكن على نحوٍ مستمر، أما إذا كان α كبيراً فإن التعليم يمكن أن يكون أسرع، ولكن سيعطي أوزاناً مهتزة حول القيمة الضرورية لإعطاء خرج صحيح لكل نماذج التدريب.

تعطى المخارج y_i للشبكة بتابع التفعيل f التالي (القفزة الواحدية):

$$y_j = f(net_j) = \begin{cases} 1 & if\ net_j > 0 \\ 0 & if\ net_j \le 0 \end{cases} \qquad (32.4)$$

حيث أصبح التابع *net* للعصبون مألوفاً لدينا ويعطى بالتركيب الخطي التالي:

$$net_j = \sum_{i=1}^{n} x_i w_{ij} \qquad (33.4)$$

في النهاية، نود عرض بعض التغيرات التـــي أجريت على شبكة بيرسبترون بسيط وهي كالآتي:

1. يمكن أن تكون مداخل الشبكة بقيم حقيقية أو ثنائية أو ثنائية القطبية.
2. قد تكون المخارج حقيقية أو ثنائية أو ثنائية القطبية.
3. يمكن أن يكون لوحدات الخرج مداخل انحياز بقيمة +1 ولكن بأوزان قابلة للتعديل.
4. قد يكون لتوابع التفعيل عتبة θ تمثل إزاحة مكانية عن المبدأ، وبالتالي:

$$net_j = \sum_{i=1}^{n} x_i w_{ij} - \theta \qquad (34.4)$$

وسنقوم بتوضيح كل هذه التغيرات في الأمثلة اللاحقة.

2.5.4 نظرية تقارب قاعدة تعليم البيرسبترون

تعتبر نظرية تقارب خوارزمية تعليم البيرسبترون من الإنجازات الهامة للباحث Rosenblatt. تنص النظرية على أن خوارزمية التعليم ستجد مصفوفة الأوزان التـــي تصنِّف تصنيفاً صحيحاً نماذج التدريب (خطياً) في عدد محدود من عمليات تكرار الأدوار (كل دور يمثل عملية التدريب لجميع نماذج دخل التدريب) عندما تصبح الشبكة مدربة.

ناقش هذه النظرية عدة باحثين وأجروا عليها بعض التعديل، لذا سنعطي برهانها المعتمد في مصادر عديدة (Hertz &Krogh & Palmer عام 1991 [23] وMinsky & Papert عام 1988[24] وArbib عام 1987[25]).

تكمن الحقيقة في أن شعاع الوزن العمودي على مستوى فصل نماذج الدخل عند كل خطوة من عمليات التعليم يمكن أن يستغل لتفسير درجة صعوبة تدريب البيرسبترون في أنواع مختلفة من الدخل.

بعبارة أخرى، إذا كان هناك شعاع وزن مثل $\mathbf{w}^*$ بحيث $f(\mathbf{x}(p).\mathbf{w}^*) = t(p)$ لكل أشعة دخل التدريب $\mathbf{x}$ ذات العدد p، وf تابع التفعيل، وt الخرج المنشود، فإنه في حالة أي شعاع وزن أولي $\mathbf{w}$ ستتقارب قاعدة تعليم البيرسبترون لشعاع وزن ما (ليس بالضرورة

وحيداً وليس بالضرورة هو w^*) يعطي استجابة صحيحة لكل نماذج التدريب وبعدد محدود من الخطوات (الأدوار).

يمكن برهان النظرية باعتبار أن مجموعة نماذج التدريب تتألف من جزأين:

$F^+ =$ {x التـــي تكون فيها قيمة الخرج المنشود مساوية +1}

$F^- =$ {x التـــي تكون فيها قيمة الخرج المنشود مساوية −1}

ومن ثم ستعرف مجموعة نماذج التدريب بمايلي:

$$F = F^+ \cup F^-$$

حيث $-F^- =$ {$-x$: التـــي تكون x في F^-}

وللتبسيط سنفرض في البرهان أن العتبة $\theta = 0$ ومعدل التعليم $\alpha = 1$. والمسألة هي الآن إيجاد شعاع الوزن w^* بحيث:

$x.w^* > 0$ إذا كان x في F^+

$x.w^* < 0$ إذا كان x في F^-

وهذا مكافئ لإيجاد شعاع الوزن w^* المحقق:

$x.w^* > 0$ إذا كان x في F

بافتراض أن قيم الخرج المنشود أصبحت تساوي +1 دوماً في مجموعة التدريب المعدلة.

لاحظ أن هذا يتحقق بعكس إشارة كل المركبات (بما في ذلك مركبة دخل الانحياز) لكل شعاع دخل كان الخرج المنشود فيه أصلاً يساوي −1. يحدث تعديل الأوزان في حالة الاستجابة غير الصحيحة لدخل التدريب المعطى للشبكة كما يلي:

$$w^{new} = w^{old} + x$$

سنقوم الآن بإثبات أن متتالية أشعة تدريب الدخل التـــي يحدث فيها تغير في الأوزان هي متتالية منتهية.

لاحظ أن أشعة دخل التدريب يجب أن يكون لكل منها مركبة إضافية (قيمتها +1 دائماً) مأخوذة بعين الاعتبار لوزن الانحياز. سنرمز للأوزان الابتدائية $w(0)$ وللأوزان الأولى الجديدة $w(1)$، وهكذا.

إذا كان $x(0)$ هو أول شعاع تدريب تحدث عنده أخطاء فإن:

$$\mathbf{w}(1) = \mathbf{w}(0) + \mathbf{x}(0) \qquad (\text{بافتراض أن } \mathbf{x}(0) . \mathbf{w}(0) \leq 0)$$

وإذا حدث خطأ آخر فسنرمز لشعاع التدريب بـ $\mathbf{x}(1)$؛ وسيكون $\mathbf{x}(1) = \mathbf{x}(0)$ إذا لم يحدث خطأ لأية أشعة تدريب أخرى وإلا $\mathbf{x}(1) \neq \mathbf{x}(0)$. وفي حالة أخرى

$$\mathbf{w}(2) = \mathbf{w}(1) + \mathbf{x}(1) \qquad (\text{بافتراض أن } \mathbf{x}(1) . \mathbf{w}(1) \leq 0)$$

وهكذا ستستمر العملية. عند الخطوة k من العملية مثلاً، ستتغير الأوزان إذا وفقط إذا فشل الوزن الحالي $\mathbf{w}(k)$ في إعطاء الاستجابة (الموجبة) الصحيحة لشعاع الدخل $\mathbf{x}(k)$ المطبق على الدخل.

ويجمع الأوزان المتغيرة المتكررة:

$$\mathbf{w}(k) = \mathbf{w}(0) + \mathbf{x}(0) + \mathbf{x}(1) + \mathbf{x}(2) + \ldots . + \mathbf{x}(k-1) \qquad (35.4)$$

نرى أن k لا يمكن أن تكون كيفياً رقماً كبيراً، فلا بد من إيجاد قيمة محددة لها.

ليكن $\mathbf{w}^*$ شـــعاع الوزن بحيث يكون $\mathbf{x} . \mathbf{w}^* > 0$ لكل أشـــعة التدريب في F، وليكن $m = \min\{\mathbf{x}. \mathbf{w}^*\}$ القيمة الصغرى للجداء الشعاعي عبر كل أشعة التدريب في F؛ هذه القيمة الصغرى موجودة فقط مادام هناك أشعة تدريب محدودة العدد.

لدينا الآن:

$$\mathbf{w}(k) . \mathbf{w}^* = [\mathbf{w}(0) + \mathbf{x}(0) + \mathbf{x}(1) + \mathbf{x}(2) + \ldots . + \mathbf{x}(k-1)] . \mathbf{w}^*$$
$$\geq \mathbf{w}(0) . \mathbf{w}^* + km$$

باعتبار $\mathbf{x}(i) . \mathbf{w}^* \geq m$ لكل $1 \leq I \leq P$.

تنص متراجحة كوشي- شـــوارتز Cauchy-Schwartz على أنه لأي شـــعاعين a وb يمكن كتابة:

$$(a . b)^2 \leq ||a||^2 \, ||b||^2$$

أو

$$||a||^2 \geq (a . b)^2 / ||b||^2$$

ومنه

$$||\mathbf{w}(k)||^2 \geq (\mathbf{w}(k) . \mathbf{w}^*)^2 / ||\mathbf{w}^*||^2$$
$$\geq (\mathbf{w}(0) . \mathbf{w}^* + km)^2 / ||\mathbf{w}^*||^2$$

وهذا يُظهر أن مربع طول شعاع الوزن أسرع في الازدياد من k^2، حيث k هو عدد

مرات تغير الأوزان خلال عملية التدريب.

على أية حال لإثبات أن هذا الطول لا يمكن أن يزداد حتـــى اللانهاية، سنعتبر:

$$\mathbf{w}(k) = \mathbf{w}(k-1) + \mathbf{x}(k-1)$$

مع العلم أن

$$\mathbf{x}(k-1) \cdot \mathbf{w}(k-1) \le 0$$

وبالاختصار والتبسيط نجد

$$\|\mathbf{w}(k)\|^2 = \|\mathbf{w}(k-1)\|^2 + 2\,\mathbf{x}(k-1) \cdot \mathbf{w}(k-1) + \|\mathbf{x}(k-1)\|^2 \le \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}(k-1)\|^2$$

وإذا كان {لكل $\mathbf{x}$ في مجموعة التدريب $\|\mathbf{x}\|^2$} $M = \max$؛ فإن:

$$\|\mathbf{w}(k)\|^2 \le \|\mathbf{w}(k-1)\|^2 + \|\mathbf{x}(k-1)\|$$

$$\le \|\mathbf{w}(k-2)\|^2 + \|\mathbf{x}(k-2)\|^2 + \|\mathbf{x}(k-1)\|^2$$

....

$$\le \|\mathbf{w}(0)\|^2 + \|\mathbf{x}(0)\|^2 + \ldots + \|\mathbf{x}(k-1)\|^2 \le \|\mathbf{w}(0)\|^2 + kM$$

وهكذا، سيزداد مربع الطول بسرعة أقل منه في حالة العلاقة الخطية مع k. وبضم المتراجحتين التاليتين نحصل على:

$$\|\mathbf{w}(k)\|^2 \ge (\mathbf{w}(0) \cdot \mathbf{w}^* + km)^2 / \|\mathbf{w}^*\|^2$$

$$\|\mathbf{w}(k)\|^2 \le \|\mathbf{w}(0)\|^2 + kM \qquad (37.4)$$

نرى بأن عدد مرات تكرار تغير الأوزان يمكن أن تكون محددة (محاطة). وبوجه خاص

$$\{(\mathbf{w}(0) \cdot \mathbf{w}^* + km)^2 / \|\mathbf{w}^*\|^2\} \qquad (37.4)$$

$$\le \|\mathbf{w}(k)\|^2 \le \{\|\mathbf{w}(0)\|^2 + kM\}$$

مرة ثانية، وبغية التبسيط الجبري، نفرض أن $\mathbf{w}(0) = 0$، فسيكون العدد الأعظمي لمرات تكرار تغير الأوزان خلال عملية التعليم ككل معطى بالعلاقة:

$$(km)^2 / \|\mathbf{w}^*\|^2 \le kM \qquad (39.4)$$

أو

$$M\,\|\mathbf{w}^*\|^2 / m^2 \ge k$$

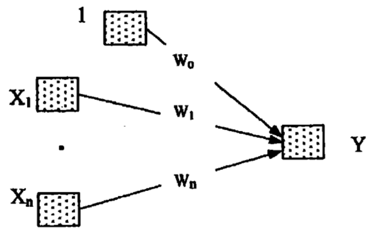
والعدد الأعظمي لمرات التكرار في عملية التعليم ككل يساوي M / m^2.

نستنتج من البرهان السابق أنه من الممكن إجراء تغيرات عديدة في قاعدة تعليم

البيرسبترون. وقد ذكر العديد من هذه التغيرات في الفصل الحادي عشـــر من كتاب Minsky وPapert عام 1988[24].

3.5.4 تطبيقات البيرسبترون في التصنيف Classification

سنعتبر الشعاعَ الثنائي الخرج في وحدات الطبقة المرافقة R في شبكة البيرسبترون البسيطة إشارة دخل لوحدة طبقة الخرج (الاستجابة) في الأمثلة التالية، وباعتبار أن الأوزان على الوصلات بين الطبقة المرافقة ووحدة الخرج هي التـــي ستعدل قيمتها فقط، لذا سنعتبر فقط الجزء الأحادي الطبقة من الشبكة الأساسية لتنفيذ المسائل التالية، حيث ستكون وظيفة الوحدات R كوحدات دخل. البنية المعتمدة موضحة في (الشكل 26.4).



الشكل 26.4: شبكة بيرسبترون لتنفيذ تصنيف مفرد

مثال 14:

استخدام البيرسبترون لتنفيذ التابع AND، بمداخل ثنائية ومخارج ثنائية القطبية، سنرى الآن كيف يتم تنفيذ التابع AND بواسطة بيرسبترون بسيط، وسنعتبر التابع المنطقي بدخل ثنائي وخرج ثنائي القطبية وسنطبق قاعدة تعليم البيرسبترون. سنعتمد معطيات التدريب المعطاة في المثال رقم 10 (الفقرة 1.4.4) في قاعدة Hebb. من الضروري وجود وزن انحياز معدل لتتمكن هذه الشبكة الأحادية الطبقة من حل هذه المسألة. بهدف التبسيط سنأخذ $\alpha = 1$ وسنجعل قيم الأوزان الأولية أصفاراً وعدد المداخل $n = 2$ كما هو مبين في (الشكل 26.4).

ولتوضيح دور العتبة سنفرض قيمتها ليست صفرية إنما ثابتة تساوي $\theta = 0.2$.

سيكون تغير الأوزان مساوياً $\Delta w = t(x_1, x_2, 1)$ إذا حدث خطأ، وصفراً في ذلك. أما استجابة وحدة الخرج فستعطى بتابع التفعيل التالي:

$$y_j = f(net_j) = \begin{cases} 1 & if\ net_j > \theta \\ 0 & if\ -\theta \le net_j \le \theta \\ -1 & if\ net_j < -\theta \end{cases} \qquad (40.4)$$

في حالة هذا التابع سيكون هناك شريط عدم القرار (undecided band) بعرض ثابت يساوي θ يفصل منطقة الاستجابة الموجبة عن منطقة الاستجابة السالبة. لذا فإن فكرة تغير الانحياز لن تطبق هنا، لأن تغير قيمة θ سيؤدي إلى تغير عرض هذا الشريط وليس مكانه.

ونلاحظ أنه عوضاً عن خط الفصل الواحد سيكون لدينا خط فصل منطقة الاستجابة الموجبة عن منطقة الاستجابة الصفرية معادلته هي:

$$w_1x_1 + w_2x_2 + w_0 > \theta \qquad (41.4)$$

ولدينا أيضاً خط فصل منطقة الاستجابة الصفرية عن منطقة الاستجابة السالبة معادلته هي:

$$w_1x_1 + w_2x_2 + w_0 < -\theta \qquad (42.4)$$

نلاحظ أيضاً أن أوزان وصلات عصبونات الدخل الفعالة هي التي يجري تحديثها $(x_i > 0)$، وستحدث أيضاً الأوزان فقط للنماذج التي لا تعطي قيمة صحيحة على الخرج.

هذا يعني أنه كلما كانت نماذج التدريب التي تعطي استجابة صحيحة أكثر حدث تعليم أقل. وهذا معاكس تماماً لتدريب وحدات ADALINE التي ستوصف فيما بعد، حيث يعتمد التعليم على الفرق بين خرج الشبكة المحسوب والخرج المنشود.

ستعطى النتائج في هذا المثال لكل دور تدريب، حيث ستتكرر عملية التعليم حتى لا يحدث أي تغير في الأوزان، عندها سيكون خرج الشبكة المحسوب مساوياً الخرج المنشود، أي تكون الشبكة قد تدربت.

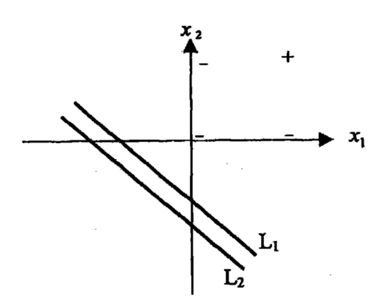
ستكون النتائج في الدور الأول كما يلي:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	0	0
									قيم أولية		
1	**1**	**1**	**0**	**0**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
1	**0**	**1**	**2**	**1**	**−1**	**−1**	**0**	**−1**	**0**	**1**	**0**
0	**1**	**1**	**1**	**1**	**−1**	**0**	**−1**	**−1**	**0**	**0**	**−1**
0	**0**	**1**	**−1**	**−1**	**−1**	**0**	**0**	**0**	**0**	**0**	**−1**

في حالة نموذج التدريب الأول (1 ,1, 1) ستكون معادلات خطوط الفصل:

$$x_1 + x_2 + 1 = 0.2 \quad (L_1)$$
$$x_1 + x_2 + 1 = -0.2 \quad (L_2)$$

مناطق استجابة الشبكة موضحة في (الشكل 4-27)، حيث يظهر أن استجابة الشبكة صحيحة في حالة أول نموذج دخل.



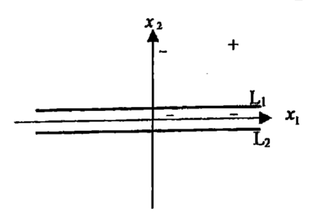
الشكل 27.4: مناطق الاستجابة لأول زوج تدريب

وفي حالة زوج التدريب الثانـــي (1, 0, 1) تكون معادلات خطوط الفصل هي:

$$x_2 = 0.2 \quad (L_1)$$
$$x_2 = -0.2 \quad (L_2)$$

ومناطق استجابة الشبكة موضحة في (الشكل 28.4)، حيث يظهر أن استجابة الشبكة

تبقى صحيحة لأول نموذج دخل فقط.



الشكل 28.4: مناطق الاستجابة من بعد زوج التدريب

وفي حالة زوج التدريب الثالث (0, 1, 1) ، وباعتبار مركبات نماذج الدخل ليست سالبة ومركبات شعاع الوزن ليست موجبة، فإن استجابة الشبكة ستكون سالبة أو مساوية للصفر.

أما في حالة الدور الرابع (0, 0, 1) فإن استجابة الشبكة ستكون سالبة لكل نماذج الدخل، لكن باعتبار أن الاستجابة في النموذج الأول (1, 1, 1) ليست صحيحة فإننا لم ننته من التدريب بعد.

لذا علينا إجراء دور ثانٍ من التدريب للأوزان الناتجة من الدور الأول كقيم أولية، وهكذا تتكرر إجراءات التدريب دوراً بعد دور لجميع نماذج التدريب حتى نحصل على استجابة صحيحة لكل نماذج التدريب.

وستكون النتائج في الدور الثاني كما يلي:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	0	−1
									قيم أولية		
1	1	1	−1	−1	1	1	1	1	1	1	0
1	0	1	1	1	−1	−1	0	−1	0	1	−1
0	1	1	0	0	−1	0	−1	−1	0	0	−2
0	0	1	−2	−1	−1	0	0	0	0	0	−2

وفي الدور الثالث:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	0	−2
									قيم أولية		
1	1	1	−2	−1	1	1	1	1	1	1	−1
1	0	1	0	0	−1	−1	0	−1	0	1	−2
0	1	1	−1	−1	−1	0	0	0	0	1	−2
0	0	1	−1	−1	−1	0	0	0	0	1	−2

وفي الدور الرابع:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	0	−2
									قيم أولية		
1	1	1	−1	−1	1	1	1	1	1	2	−1
1	0	1	0	0	−1	−1	0	−1	0	2	−2
0	1	1	0	0	−1	0	−1	−1	0	1	−3
0	0	1	−3	−1	−1	0	0	0	0	1	−3

وفي الدور الخامس:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	1	−3
									قيم أولية		
1	1	1	−1	−1	1	1	1	1	2	2	−2
1	0	1	0	1	−1	−1	0	−1	1	2	−3
0	1	1	−1	−1	−1	0	0	0	1	2	−3
0	0	1	−3	−1	−1	0	0	0	1	2	−3

وفي الدور السادس:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	1	1	−3
									قيم أولية		
1	1	1	−1	−1	1	1	1	1	2	2	−2
1	0	1	0	0	−1	−1	0	−1	1	2	−3
0	1	1	−1	−1	−1	0	0	0	1	2	−3
0	0	1	−3	−1	−1	0	0	0	1	2	−3

وفي الدور السابع:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	1	2	−3
									قيم أولية		
1	1	1	0	0	1	1	1	1	2	3	−2
1	0	1	0	0	−1	−1	0	−1	1	3	−3
0	1	1	0	0	−1	0	−1	−1	1	2	−4
0	0	1	−4	−1	−1	0	0	0	1	2	−4

وفي الدور الثامن:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	1	2	−4
									قيم أولية		
1	1	1	−1	−1	1	1	1	1	2	3	−3
1	0	1	−1	−1	−1	0	0	0	2	3	−3
0	1	1	0	0	−1	0	−1	−1	2	2	−4
0	0	1	−4	−1	−1	0	0	0	2	2	−4

وفي الدور التاسع:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	2	2	–4
									قيم أولية		
1	1	1	0	0	1	1	1	1	3	3	–3
1	0	1	0	0	–1	–1	0	–1	2	3	–4
0	1	1	–1	–1	–1	0	0	0	2	3	–4
0	0	1	–4	–1	–1	0	0	0	2	3	–4

وفي الدور العاشر:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	2	3	–4
									قيم أولية		
1	1	1	0	0	1	0	0	0	2	3	–4
1	0	1	–2	–1	–1	0	0	0	2	3	–4
0	1	1	–1	–1	–1	0	0	0	2	3	–4
0	0	1	–4	–1	–1	0	0	0	2	3	–4

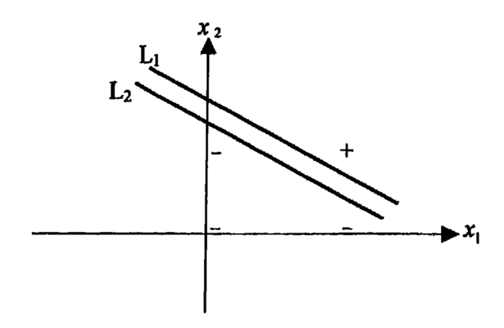
وباعتبار أن قيمة تغير الأوزان أصبحت تساوي الصفر في الدور العاشر، فإن الشبكة أصبحت مدربة تدريباً كاملاً، وهكذا تعطى الاستجابة الموجبة لجميع النقاط:

وهي معادلة المستقيم (L_1) $\quad 2x_1 + 3x_2 - 4 > 0.2 \quad \Rightarrow \quad x_2 = -\frac{2}{3}x_1 + \frac{7}{5}$

والاستجابة السالبة تعطى لجميع النقاط:

وهي معادلة المستقيم (L_2) $\quad 2x_1 + 3x_2 - 4 < -0.2 \quad \Rightarrow \quad x_2 = -\frac{2}{3}x_1 + \frac{19}{15}$

ومناطق الفصل موضحة في (الشكل 29.4).



الشكل 29.4: الاستجابة النهائية للتابع AND بقاعدة تعليم البيرسبترون

مثال 15:

استخدام البيرسبترون لتنفيذ التابع **OR**، بمداخل ثنائية ومخارج ثنائية القطبية يمكن حساب التابع المنطقي OR في حالة مداخل ثنائية ومخرج ثنائي القطبية، حيث $\alpha=0$ ، $\theta=0.2$.

وذلك بإجراء نفس خطوات الخوارزمية السابقة حتى يتم تدريب الشبكة والحصول على تغير أوزان مساو للصفر؛ وهذا يعني أن الخرج المحسوب يساوي الخرج المنشود.

وستكون النتائج كما يلي:

المداخل الخارجية			الدخل التركيبي	الخرج المحسوب	الخرج المنشود	تغير الأوزان			الأوزان		
									w_1	w_2	w_0
x_1	x_2	x_0	net	y	t	Δw_1	Δw_2	Δw_0	0	0	0
									قيم أولية		
1	1	1	0	0	1	1	1	1	1	1	1
1	0	1	2	1	1	0	0	0	1	1	1
0	1	1	2	1	1	0	0	0	1	1	1
0	0	1	1	1	−1	0	0	−1	1	1	0

1	1	1	2	1	1	0	0	0	1	1	0
1	0	1	1	1	1	0	0	0	1	1	0
0	1	1	1	1	1	0	0	0	1	1	0
0	0	1	0	0	–1	0	0	–1	1	1	–1
1	1	1	1	1	1	0	0	0	1	1	–1
1	0	1	0	0	1	1	0	1	2	1	0
0	1	1	1	1	1	0	0	0	2	1	0
0	0	1	0	0	–1	0	0	–1	2	1	–1
1	1	1	2	1	1	0	0	0	1	1	–1
1	0	1	1	1	1	0	0	0	2	1	–1
0	1	1	0	0	1	0	1	1	2	2	0
0	0	1	0	0	–1	0	0	–1	2	2	–1
1	1	1	3	1	1	0	0	0	2	2	–1
1	0	1	1	1	1	0	0	0	2	2	–1
0	1	1	1	1	1	0	0	0	2	0	–1
0	0	1	–1	–1	–1	0	0	–1	2	0	–1

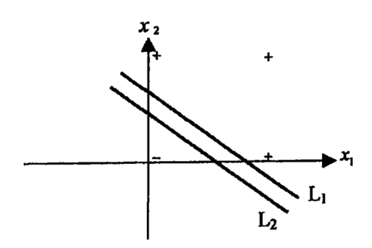
من هذه النتائج نلاحظ أن الشبكة أصبحت مدربة في الدور الخامس. تعطى الاستجابة الموجبة لجميع النقاط المحققة للمعادلة:

وهي معادلة المستقيم (L_1) $\quad 2x_1 + 2x_2 - 1 > 0.2 \;\Rightarrow\; x_2 = -x_1 + \frac{3}{5}$

والاستجابة السالبة تعطى لجميع النقاط:

وهي معادلة المستقيم (L_2) $\quad 2x_1 + 2x_2 - 1 < -0.2 \;\Rightarrow\; x_2 = -x_1 + \frac{2}{5}$

ومناطق الفصل موضحة في (الشكل 30.4).



الشكل 30.4: الاستجابة النهائية للتابع **OR** بقاعدة تعليم البيرسبترون

4.5.4 تطبيقات البيرسبترون في تعرف الأحرف

في هذا التطبيق سنستخدم بيرسبترون لتمييز الأحرف، وسيجري تنفيذ هذا التطبيق على النحو التالي:

1. طبق خوارزمية التدريب على مجموعة الأوزان
2. لكي يكون كل شعاع دخل **x** مصنفاً علينا تطبيق الخطوات 4 و3
3. ضع تفعيلات وحدات الدخل
4. احسب استجابة وحدة الخرج:

$$y = f(net) = \begin{cases} 1 & if\ net > \theta \\ 0 & if\ -\theta \le net \le \theta \\ -1 & if\ net < -\theta \end{cases}$$

$$net = \sum_i x_i w_i$$

مثال 16:

استخدام شبكة البيرسبترون لتصنيف الأحرف من تشكيلات مختلفة، مع صف خرج وحيد في هذا المثال سنستخدم شبكة بيرسبترون لتصنيف أحرف من تشكيلات مختلفة، مع صف خرج وحيد. سنعتمد 21 نموذج دخل موضحة في (الشكل 31.4) كأمثلة عن الحرف A أو "ليس A". لسهولة التمثيل أعطينا كل نقطة (pixel) غير مضاءة من الحرف الرمز "0"، وكل نقطة مضاءة الرمز " # ".

بعبارة أخرى سندرب الشبكة على تصنيف كل هذه الأشعة هل تنتمي إلى الصف A أو لا.

ستكون قيمة الخرج من أجل كل نموذج إما +1 وإما 1−.

الدخل من التشكيلة الأولى:

```
00##000   ######0   00#####   #####00   #######   000####   ###00##
000#000   0#0000#   0#0000#   0#000#0   0#0000#   00000#0   0#00#00
000#000   0#0000#   #000000   0#0000#   0#00000   00000#0   0#0#000
00#0#00   0#0000#   #000000   0#0000#   0#0#000   00000#0   0##0000
00#0#00   0#####0   #000000   0#0000#   0###000   00000#0   0##0000
0#####0   0#0000#   #000000   0#0000#   0#0#000   00000#0   0#0#000
0#000#0   0#0000#   #000000   0#0000#   0#00000   0#000#0   0#00#00
0#000#0   0#0000#   0#0000#   0#000#0   0#0000#   0#000#0   0#000#0
###0###   ######0   00####0   #####00   #######   00###00   ###00##
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثانية :

```
000#000   ######0   00###00   #####00   #######   00000#0   #0000#0
000#000   #00000#   0#000#0   #0000#0   #000000   00000#0   #000#00
000#000   #00000#   #00000#   #00000#   #000000   00000#0   #00#000
00#0#00   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
00#0#00   ######0   #000000   #00000#   #####00   00000#0   ##00000
0#000#0   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
0#####0   #00000#   #00000#   #00000#   #000000   0#000#0   #00#000
0#000#0   #00000#   0#000#0   #0000#0   #000000   0#000#0   #000#00
0#000#0   ######0   00###00   #####00   #######   00###00   #0000#0
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثالثة :

```
000#000   ######0   00###0#   #####00   #######   0000###   ###00##
000#000   0#0000#   0#000##   0#000#0   0#0000#   00000#0   0#000#0
00#0#00   0#0000#   #00000#   0#0000#   0#00#00   00000#0   0#00#00
00#0#00   0#0000#   #000000   0#0000#   0####00   00000#0   0#0#000
0#000#0   0#####0   #000000   0#0000#   0#00#00   00000#0   0##0000
0#####0   0#0000#   #000000   0#0000#   0#00000   00000#0   0#0#000
#00000#   0#0000#   #00000#   0#0000#   0#00000   00000#0   0#00#00
#00000#   0#0000#   0#000#0   0#000#0   0#0000#   0#000#0   0#000#0
##000##   ######0   00###00   #####00   #######   00###00   ###00##
A......   .B.....   ..C....   ...D...   ....E..   .....J.   ......K
```

الشكل 31.4: نماذج دخل التدريب والخرج المنشود

الشبكة التـــي ستنفذ هذا التطبيق موضحة في (الشكل 26.4) في حالة $n = 7 \times 9 = 63$.
سيكون هناك ثلاث عينات عن A و18 أنموذجاً عن "ليس A" كما هو موضح في

(الشكل 31.4). بالطبع نستطيع استعمال نفس الأشعة كأمثلة عن B "وليس B" والتدريب سيتم بنفس الإجراءات. وبسبب استعمال طبقة واحدة في الشبكة، فإن الأوزان لوحدة الخرج المخصصة للحرف A ليس لها أي تقاطع (علاقة) مع الأوزان المخصصة للحرف B. لذا نستطيع حل هاتين المسألتين بوقت واحد بالسماح لشعاع الأوزان لكل وحدة خرج.

سيكون للشبكة 9 × 7 = 63 وحدة دخل ووحدتـــي خرج؛ الأولى من أجل الحرف A أو "ليس A" والثانية للحرف B أو "ليس B". باستخدام هذه الفكرة نستطيع تشكيل شبكة لها سبع وحدات خرج، واحدة لكل حرف من التشكيلات السبعة التـــي نرغب بتدريب الشبكة عليها.

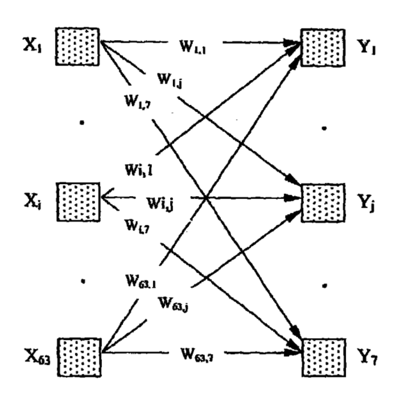
مثال 17:

استخدام شبكة البيرسبترون لتصنيف الأحرف من تشكيلات مختلفة، مع صفوف خرج متعددة.

في هذه الحالة كل شعاع دخل يحوي 7 × 9 = 63 مركبة، كل منها يمثل حرف عُبِّر عنه كنموذج ثنائي على شبكة مؤلفة من 7 × 9 نقطة (عُبِّر عن النقطة التـــي ليست مضاءة (ليست من الحرف) بالرمز "0"، أما النقاط المضاءة (المنتمية للحرف) عُبِّر عنها بالرمز"#"). نماذج التدريب موضحة في (الشكل 31.4).

هناك سبع فئات (أحرف) ينتمي شعاع الدخل إليها، ومن ثم سيكون لدينا سبع وحدات خرج، وشعاع الخرج يتألف من ســـبع مركبات كل منها يمثل حرفاً من الأحرف الســـبعة A ,B ,C ,D ,E ,J ,H، ولسهولة القراءة سيكون ترتيب نموذج الخرج المنشود إذا كان الدخل هو الحرف A كمـــا يلـــي: (A . . . . . .)، وإذا كان الدخل هو الحـــرف B كمـــا يلـــي: (. B . . . . .)، وهكذا في بقية الأحرف على الخرج. الشبكة موضحة في (الشكل 32.4).

بالطبع يجب أن تحول نماذج تدريب الدخل والاستجابات المرغوب فيها إلى شكل مناسب لعمل الشبكة العصبونية الصنعية. سنختار التمثيل ثنائي القطبية لأنه أفضل حسابياً من التمثيل الثنائي. وبالتالي جميع نماذج الدخل يجب أن تحول إلى أشعة ثنائية القطبية كما هو مشروح في الفقرة (2.4.4) في المثال 12.



الشكل 32.4: شبكة بيرسبترون لتصنيف الدخل إلى سبعة أحرف

مثلاً، نموذج الخرج المنشـود (A . . . . . .) يصبح بعـد التحويل إلـى ثنائي القطبيـة (1- ,1- ,1- ,1- ,1- ,1- ,1) والنموذج المنشود (B . . . . .) سيمثل بالشعاع ثنائي القطبية التالي: (1- ,1- ,1- ,1- ,1- ,1 ,1-)، وهكذا. نماذج دخل التدريب والاستجابات المنشودة موضحة في (الشكل 33.4).

ستصبح خوارزمية التدريب المعدلة من أجل سبع فئات خرج، مع عتبة صفرية ومعدل تعليم قيمته تساوي الواحد، وباستعمال التمثيل ثنائي القطبية كما يلي:

1. إعطاء الأوزان والانحيازات قيماً أولية (صفر أو قيم كيفية صغيرة).
2. مادام شرط توقف العمل غير محقق، نفذ الخطوات من 3 إلى 7.
3. في حالة كل زوج تدريب ثنائي القطبية (s, t) نفذ الخطوات من 4 إلى 6.
4. ضع تفعيل كل وحدة دخل؛

$$x_i = s_i \quad , \quad i = 1,2,\ldots,n$$

5. احسب تفعيل كل وحدة خرج؛

$$net_j = w_{0j} + \sum_i x_i w_{ij}$$

$$j = 1,2,\ldots,m \quad , y_j = f(net_j) = \begin{cases} 1 & if\ net_j > \theta \\ 0 & if\ -\theta \le net_j \le \theta \\ -1 & if\ net_j < -\theta \end{cases}$$

6. تحديث الأوزان والانحياز

إذا كان $t_j \ne y_j$ فإن

$$w_{0j}^{new} = w_{0j}^{old} + t_j \quad , \quad j = 1,2,\ldots,m$$

$$w_{ij}^{new} = w_{ij}^{old} + t_j x_i \quad , \quad i = 1,2,\ldots,n$$

وإلا

$$w_{0j}^{new} = w_{0j}^{old} \quad , \quad j = 1,2,\ldots,m$$

$$w_{ij}^{new} = w_{ij}^{old} \quad , \quad i = 1,2,\ldots,n$$

7. اختبر شرط توقف العمل:

إذا كان $\Delta w_{ij} = 0$ توقف، وإلا تابع العمل.

استطاعت الشبكة بعد تدريبها تصنيف كل أشعة التدريب تصنيفاً صحيحاً.

من (الشكل 33.4) نلاحظ أن كل واحد من نماذج الدخل هو نموذج تدريب الدخل المعتمد مع قليل من النقاط المتغيرة (الضجيج). يشار إلى النقاط الضجيجية المعبرة عن الاختلاف بين نموذج الدخل ونموذج التدريب بوضع "*" لكل نقطة هي "on" (واحد) في نموذج الدخل الضجيجي ولكن كانت " off" (صفر) في نموذج التدريب، و"x" لكل نقطة هي " off" ولكن كانت "on".

5.5.4 قاعدة تعليم المحفظة

The Pocket Learning Algorithm

لقد تعلمنا درساً هاماً من خوارزمية تعليم البيرسبترون هو أن عملية التعليم يجب أن تكون متواصلة بشكل حثيث. عندما توجد مجموعة الأوزان التـي تعطي تصنيفاً صحيحاً يجب تنفيذ إجراء ما لمكافئة الشبكة أيضاً. إن عقاب الشبكة على أخطائها فقط يعتبر لوناً من ألوان التقوية السلبية.

الدخل من التشكيلة الأولى :

```
00##000   ###*##0   00#####   ####*00   ##x##*#   000##x#   #x#0*0#
00*#000   0x0000#   *#00*0#   0#0*0#0   0#0000#   00000#0   0#00#00
000#000   0#0000#   #000000   0#0000x   0#00000   00000#0   0#0#000
00#0#*0   0#*000#   #*00000   0#*000#   0#0#000   00000#0   0##0000
00#0#00   0####*0   x000000   0#0000x   0#x#*00   00000#0   0##0000
0####0    0#0*00#   #000000   0#0000#   0#0#000   00000x0   0#0#000
0#0*0#0   0#*000#   #000000   0#0000#   0#00000   0x000#0   0#00#*0
0#000#0   0#0000#   0#00*0#   0#*00#0   *#0000#   0x00*#0   0#*00#0
#x#0#x#   #####0    00####0   ####*00   #######   00###00   ##x00##
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثانية :

```
000#000   #####0    00###00   x####00   x######   00000#0   x0000x0
000x000   #00000#   0#000#0   #0000#0   #000000   00000#0   #000#00
000#000   x00000#   #00000#   #00000#   #000000   00000#0   #00#000
00#0#00   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
00#0x00   x##x##0   #*0*0*0   #**000#   x####00   00000#0   x#00000
0#000#0   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
0###x#0   #00000#   #00000x   x00000#   #0*0000   0x000#0   #00#000
*x000x0   #00*00x   0#000x0   #0000#0   #0*0000   *x000#0   #000#00
*x00*x*   ######0   00###00   x####00   x######   *0x##00   x0000x0
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثالثة :

```
000#000   ######0   00###0#   ####00    #######   0000###   x##00#x
000#000   *#0000#   0#000##   0#000#0   0#0000#   00000*0   0#000#0
00#0#00   0#0000#   #00000#   0#0000#   0#00#00   00000#*   0#00#00
00#0#00   0#0000#   #000000   0#0000#   *x##x00   00000#0   0#0x000
0#000#0   *####*0   #00000*   0**000x   0#00#00   00000x0   0##0000
0x#x#x0   0#0000#   x000000   0#0000#   0#00000   00000#0   0#0#000
#00000#   *#0000x   #00000#   0#0000x   0#00000   00000#0   0#00#00
#00000#   0#0000#   0#000#0   0#*00#0   0#0000#   0#000x0   0#000x0
##**0##   ######0   00x##00   #####00   x###x##   00###00   x##00##
A......   .B.....   ..C....   ...D...   ....E..   .....J.   ......K
```

الشكل 33.4: نماذج دخل ضجيجية باستعمال شبكة بيرسبترون

من الواضح أن الشكل الحالي لخوارزمية تعليم البيرسبترون لا يمكن استخدامه مطلقاً في مسائل الفصل غير الخطية، إذ عوضاً عن إيجاد مجموعة من الأوزان قادرة على التصنيف الجيد فإن الخوارزمية ستشرد أو تتيه، ولكن ليس بالضرورة من أجل كل نماذج التدريب سيحدث هذا الشرود، وتقوم باختيار مجموعة من الأوزان ضعيفة التصنيف؛ وبالتالي معظم الأمثلة أو الأشكال لن تصنف جيداً.

هذا النوع من السلوك هو الذي دفع Gallant عام 1993[22] ليسمي خوارزمية تعليم البيرسبترون بضعيفة السلوك. والسبب الرئيس لهذا الضعف هو أن الخوارزمية لا تتعلم عندما يكون إنجازها جيداً، لأن عملية التعليم لا تكافئ الإنجاز القوي وتعاقب فقط الإنجاز الضعيف.

إحدى الطرق المقترحة للتغلب على هذه المشكلة هي تعديل الخوارزمية لكي تقوم بالاحتفاظ بنسخة منفصلة عن مجموعة الأوزان **w** خلال التدريب، ومتـــى كان لمجموعة الأوزان الجديدة جريان أطول للتصانيف الصحيحة من المجموعة المخبأة، تخزَّن هذه المجموعة الجديدة محل القديمة المخزنة سابقاً. أطلق Gallant عام 1993 على هذا التعديل لخوارزمية تعليم البيرسبترون اسم خوارزمية المحفظة (pocket) لأنها تحتفظ بنسخة عن الأوزان ذات التصنيف الأفضل لاستعمالها لاحقاً.

هناك نسخة معدلة عن هذه الخوارزمية اسمها خوارزمية المحفظة rachet، وهي تحسين للخوارزمية الأساسية في حالتـــي التنفيذ الحالي للتصانيف الصحيحة الذي يجب أن يكون أطول جرياناً من خوارزمية المحفظة، وللأوزان الجديدة التـــي يجب أن تصنف تصنيفاً صحيحاً أشكالاً أكثر من مجموعة أوزان المحفظة. أظهرت خوارزمية المحفظة أنها تعطي مجموعات وزن جيدة في مسائل التصنيف، بعد عدد محدود من نماذج التدريب.

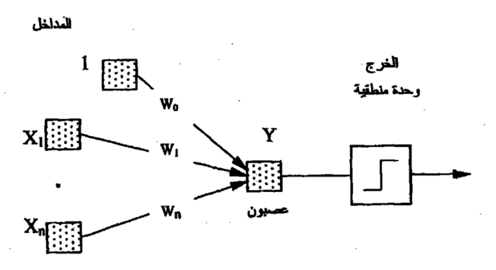
أثبتت نظرية التقارب المبرهنة من قبل Gallant أن أوزان المحفظة تصبح أمثلية باحتمال عالي بعد عدد محدود من التكرارات. عرفت الأمثلية عندما صنَّفت مجموعة الأوزان تصنيفاً صحيحاً أعظم عدد ممكن من الأشكال (في حالة الفصل الخطي). وقد أجريت دراسات مقارنة على ثلاثة مسائل للتصنيف، جميعها أثبتت أن خوارزمية المحفظة مع rachet أقوى من خوارزمية البيرسبترون في إنجاز التصنيف والاستقرار.

لقد استعملت شبكات البيرسبترون البسيطة في تطبيقات عديدة مثل مهام تعرّف الأشكال كمعالجة إشارة الكلام من قبل Rosenblatt عام 1962[26] وBurr عام 1988[27]، وكذلك استعملت في مهام تصنيف الأنظمة الخبيرة من قبل Gallant عام 1993[22]، واستعملت في قضايا التنبؤ، وفي الرؤية الحاسوبية من قبل Kollias وKuppa عام 1988[28].

6.4 الشبكات العصبونية الصنعية ADALINE

اقترحت شبكات ADALINE (ADAptive LInear NEuron) من قبل idrow & Hoff عام 1960[29]. وتستعمل هذه الشبكات عموماً تفعيلات ثنائية القطبية (+1, -1) كقيم لمداخلها وخرجها المنشود (ولكنها ليست ملزمة بمثل هذا التمثيل، ويمكن أن تكون المعطيات حقيقية أو ثنائية). إن أوزان الوصلات من وحدات الدخل إلى وحدة ADALINE قابلة للتعديل، إضافة إلى ذلك فإن لشبكة ADALINE انحيازاً يعمل كوزن قابل للتعديل على وصلة لها تفعيل مساوٍ للواحد دائماً. يوضح (الشكل 4-34) بنية هذا الشبكة.

بوجه عام، دُرِّبت ADALINE باستعمال قاعدة دلتا المعروفة بمتوسط المربعات الصغرى LMS (Least mean squares) أو قاعدة Widrow-Hoff. يمكن أن تُستعمل هذه القاعدة في الشبكات وحيدة الطبقة بوحدات خرج عديدة؛ في هذه الحالة يعتبر ADALINE حالة خاصة، لأن هناك وحدة واحدة في الخرج.



الشكل 34.4: شبكة ADALINE

خلال التعليم، يكون تفعيل الوحدة مساوياً لدخلها net؛ وهذا يعني أن تابع التفعيل

لوحدة الخرج هو تابع خطي $f(net) = net$.

تبحث قاعدة التعليم عن أصغر متوسط مربع الخطأ بين قيمة التفعيل وقيمة الخرج المنشود. وهذا يسمح للشبكة باستمرار التعليم على كل نماذج التدريب، حتى بعد إعطاء الاستجابة الصحيحة على الخرج (إذا طبق تابع العتبة) لبعض النماذج.

بعد التدريب، إذا استُعملت الشبكة في مهام تصنيف النماذج التـــي فيها الخرج المرغوب به سيكون إما +1 وإما -1، فإن تابع العتبة سيطبق على دخل الوحدة لحصول التفعيل. إذا كان دخل (ADALINE) أكبر أو يساوي الصفر فإن تفعيلها سيكون +1 على الخرج، وإلا سيكون الخرج -1.

يمكن نمذجة أي مسألة تقتضي أن تكون نماذج الدخل الموافقة لقيمة خرج +1 مفصولة خطياً عن نماذج الدخل الموافقة لقيمة خرج -1، باستعمال شبكة أدالين. ومع أن هذه الشبكات قادرة فقط على التصنيف الخطي كما في البيرسبترون، فقد استُعملت بفعالية في تطبيقات هندسية عديدة كما سنرى فيما بعد.

1.6.4 خوارزميات تعليم Widrow-Hoff

تشبه خوارزميةُ التعليم المقترحة من قبل Widrow-Hoff عام 1960 خوارزمية تعليم البيرسبترون؛ فهي شكل من التعليم بمعلم. يمكن أن تطبق هذه الخوارزمية على أية شبكة عصبونية صنعية بطبقة وحيدة أمامية التغذية وبتوابع تفعيل تفاضلية. ففي أدالين تكون هذه التوابع خطية، أي:

$$y = \mathbf{x}^{\mathrm{T}}\,\mathbf{w} = \sum_{i} x_i w_i \qquad (43.4)$$

حيث w_i الوزن على الوصلة بين عنصر نموذج الدخل i وعصبون الخرج الوحيد (يرسل خرج العصبون الخطي إلى وحدة منطقية العتبة للحصول على خرج نهائي ثنائي القطبية).

في حالة دخل معطى **x** كنموذج تدريب، تعطى قاعدة التعليم بالصيغة التكرارية التالية:

$$w_i^{new} = w_i^{old} + \alpha\,(t - y)x_i \qquad (44.4)$$

حيث α عامل التعليم، وt قيمة الخرج المرغوب به لشعاع الدخل **x**، وy قيمة الخرج المحسوب للعصبون. بعد ذلك سيكون تعليم أدالين حالة خاصة في حال m = 1، حيث m

عدد الوحدات في طبقة خرج الشبكة.

وهكذا نرغب بتعديل الأوزان لتقليل الخطأ الكلي E_{tot} عبر كل وحدات الخرج وكل نماذج التدريب $p = 1, 2, ..., P$، حيث

$$E^p = \sum_{j=1}^{m}(t_j^p - y_j^p)^2 \quad , \quad E_{tot} = \sum_{p=1}^{P} E^p \qquad (45.4)$$

E^p هو الخطأ في نموذج مفرد p عبر كل وحدات الخرج ويعطى كمجموع مربع الأخطاء. يمكن أن يصغَّر الخطأ بتعديل الأوزان بالتناسب مع التدرج السالب؛ أي اتجاه التناقص الأسرع في تابع الخطأ E_{tot} لكل تغير في الوزن.

وهكذا نريد التعبير عن تغير الوزن Δw_{ij} تناسبياً مع التدرج السالب للخطأ؛ أي:

$$\Delta w_{ij} = -\eta \frac{\partial E_{tot}}{\partial w_{ij}} = -\eta \sum_{p=1}^{P} \frac{\partial E^p}{\partial w_{ij}} \qquad (46.4)$$

حيث μ ثابت موجب. بتعويض العلاقة (45.4) في العلاقة (46.4) وبأخذ المشتقات لكل حد في المجموع يكون لدينا:

$$\frac{\partial E^p}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\left(\sum_{j=1}^{m}\left(t_j^p - \sum_{i=1}^{n} w_{ij}\, x_i^p\right)^2\right) = -2\sum_{j=1}^{m}\left(t_j^p - y_j^p\right)x_i^p$$

للموافقة مع ما سيتبع سنهمل أدلة نماذج التدريب في العلاقة السابقة بوضع $P = 1$. ولما كان $m = 1$ (وحدة خرج وحيدة) فسنهمل الأدلة j، ومن ثم نستطيع إعطاء قاعدة دلتا لتعليم أدالين كما يلي:

$$\Delta w_i = \eta\, (t - y) x_i = \eta\, \varepsilon\, x_i \qquad (47.4)$$

حيث الضارب 2 تُضمِّن في عامل التعليم η، ويمثل ε مقدار الخطأ $(t - y)$. لاحظ التشابه الكبير بين قاعدة تعليم البيرسبترون وقاعدة تعليم دلتا.

من الجدير بالملاحظة أنه عندما تكون المداخل ثنائية و$x_i = 0$ في (47.4) لا يحدث تعليم. وهذا يوضح لماذا يكون العمل على أشعة دخل ثنائية القطبية أكثر مناسبة.

نحصل على الشكل المعياري للمعادلة (47.4) بتقسيم الطرف الأيمن على مربع نظيم الشعاع x:

$$\Delta w_i = \frac{\eta \varepsilon \ x_i}{\|\mathbf{x}\|^2}$$

من هذه العلاقة نستطيع أن نثبت أن الخطأ انخفض عند التحديث (الخطوة) رقم k إلى قيمة $-\eta\,\varepsilon_k$:

$$\begin{aligned} \Delta\varepsilon_k &= \Delta(t_k - y_k) = \Delta(t_k - \mathbf{x}_k^{\mathrm{T}}\mathbf{w}_k) = -\mathbf{x}_k^{\mathrm{T}}\Delta\mathbf{w}_k \\ &= \frac{-\eta\varepsilon_k \mathbf{x}_k^{\mathrm{T}}\mathbf{x}_k}{\|\mathbf{x}_k\|^2} = -\eta\varepsilon_k \end{aligned} \qquad (48.4)$$

باعتبار أنه في الشكل الشعاعي لدينا: $\Delta\mathbf{w}_k = \eta\varepsilon_k \frac{\mathbf{x}_k}{\|\mathbf{x}_k\|^2}$.

الإجراء السابق ــ كما هو ملاحظ ــ شكل من التدرج السالب، وسيكون التقارب في w مؤكداً باعتبار أن سطح الخطأ قطعي ناعم بقيمة صغرى وحيدة (Kohonen عام 1977[30]).

تعطينا مصفوفة الوزن ترافقاً لنماذج الدخل/الخرج على نحو كامل وذلك عندما تكون نماذج دخل التدريب مستقلة خطياً. أما عندما لا تكون المداخل مستقلة خطياً، فإن نماذج الخرج الفعلية ستختلف عن نماذج الخرج المنشودة في معنـــى المربعات الصغرى الأصغرية (كمجموع الأخطاء المربعة).

فإذا كانت نماذج الدخل مستقلة خطياً، يمكن أن تطبق التحويلات على أشعة الدخل والخرج (بافتراض وجود m عقدة خرج) للحصول على تغير في الأساس. سنطبق التحويلات للحصول على أشعة نموذج خرج ودخل متعامدة. إذا طبقت التحويلات الناتجة مشتركة على مصفوفة الوزن في قاعدة تحديث الأوزان، يمكننا عزل مصفوفة الارتباط لنماذج الدخل (Stone عام 1986[31]). في هذا الشكل الجديد، من الممكن الحصول على معالم إضافية لقاعدة دلتا.

والآن سنلخص العملية كما يلي: سنفترض أن التحويل على نماذج الدخل x يعطى بمصفوفة P، والتحويل على الخرج y يعطى بمصفوفة Q فيكون:

$$\mathbf{x}^* = \mathbf{P}\,\mathbf{x}\,,\ \mathbf{y}^* = \mathbf{Q}\,\mathbf{y}$$

ويجب أن نطبق **Q** على المخارج المرغوب بها **t** للحصول على $\mathbf{t^* = Q\,t}$. أخيراً نطبق التحويلات جميعاً على **w** فنحصل على $\mathbf{w^*\,x^* = y^*}$. من هذه المعادلة الأخيرة يمكن كتابة:

$$\mathbf{w^*\,P\,x = Q\,y} \Rightarrow \qquad \mathbf{Q^{-1}\,w^*\,P\,x = y = w\,x}$$

وباعتبار $\mathbf{Q^{-1}\,w^*\,P = w}$، فإن تحويل المصفوفة **w** سيكون:

$$\mathbf{w^* = Q\,w\,P^{-1}}$$

وبتطبيق هذا التحويل على قاعدة التعليم (المعادلة (47.4) بالشكل الشعاعي) نحصل على المعادلة المحولة:

$$\Delta \mathbf{w}^* = \eta\varepsilon^* (\mathbf{x}^*)^{\mathbf{T}}\,\mathbf{C} \qquad (49.4)$$

حيث $\varepsilon^* = t^* - y^*$.

تحتوي المصفوفة **C** في العلاقة السابقة على المعلومات الارتباطية بين نماذج الدخل الأصلية. العنصر في السطر **i** والعمود **j** من المصفوفة **C**، c_{ij}، هو جداء داخلي للنموذجين $\mathbf{x}_i$ و$\mathbf{x}_j$. وبهذا نستطيع إثبات أن الخرج الناتج لدخل معطى يمكن أن يفسر كمتوسط مثقل لنماذج الخرج المنشود. محاكاة التعليم خطوة بخطوة باستعمال الأساس الجديد تبين هذه العلاقة جيداً (Stone، 1986[31]).

عندما تكون نماذج الدخل أشعة عشوائية مستنتجة من توزيع مستقر، يمكن أن نرى بعد نهاية التعليم أن القيمة المتوقعة لحسابات الخرج ستكون مساوية للقيمة المتوسطة للخرج المنشود. وإذا كانت المداخل **x** والأهداف **t** موزعة طبيعياً بمتوسطات صفرية، فيمكن إثبات أن القيمة المتوقعة لشعاع الوزن **w** تحقق:

$$E(\mathbf{y}) = E(\mathbf{t}|\mathbf{x}) \qquad (50.4)$$

بعبارة أخرى موجزة، لكل دخل خاص، تعطي قاعدة دلتا خرجاً يساوي متوسط الخرج المنشود لهذا الدخل الخاص.

2.6.4 تطبيقات أدالين في تصنيف النماذج

بعد تدريب شبكة أدالين يمكن استعمالها لتصنيف نماذج الدخل، ستكون قيم الخرج

المنشودة ثنائية القطبية أو ثنائية، وسيطبق تابع الخطوة كتابع تفعيل لوحدة الخرج.

ستكون الخوارزمية، عموماً، كما يلي:

1. إعطاء الأوزان قيماً أولية صغيرة، وفرض قيمة عامل التعليم α.
2. مادام شرط التوقف غير محقق، كرر الخطوات من 3 إلى 7.
3. لكل زوج دخل تدريب ثنائي القطبية s;t، كرر الخطوات من 4 إلى 7.
4. ضع تفعيلات وحدات الدخل:

$$x_i = s_i \quad ,i=1,2,\ldots,n$$

حيث s_i مركبات شعاع دخل التدريب.

5. احسب دخل الشبكة net إلى وحدة الخرج:

$$net = w_0 + \sum_i x_i w_i \quad , \quad i=1,2,\ldots,n$$

6. حدث الانحياز والأوزان، $i=1,2,\ldots,n$:

$$w_0^{new} = w_0^{old} + \alpha(t - net)$$
$$w_i^{new} = w_i^{old} + \alpha(t - net)x_i$$

7. طبق تابع التفعيل:

$$y = \begin{cases} 1 & if \quad net \geq 0 \\ -1 & if \quad net < 0 \end{cases}$$

8. اختبر شرط التوقف: إذا كان أكبر تغير للأوزان أصغر من المجال المسموح به توقف، وإلا تابع تنفيذ الخوارزمية.

مثال 18:

استخدام أدالين لتنفيذ التابع **AND**، بمداخل ثنائية ومخارج ثنائية القطبية، لقد رأينا مما سبق من خوارزمية تعليم أدالين، أن الشبكة مصممة لإيجاد الأوزان بأقل خطأ ممكن:

$$E = \sum_{p=1}^{4} E^p = \sum_{p=1}^{4} (x_1^p w_1 + x_2^p w_2 + w_0 - t^p)^2$$

حيث $net = x_1^p w_1 + x_2^p w_2 + w_0$ هو دخل الشبكة إلى وحدة الخرج للنموذج p ‏(P = 4)، وt^p هو الهدف المنشود لهذا النموذج. الأوزان التـــي تجعل الخطأ أصغرياً هي

$w_0 = -3/2$، $w_1 = 1$، $w_2 = 1$. وسيكون خط الفصل هو $x_1 + x_2 - 3/2 = 0$. إن مربع الخطأ الكلي في نماذج التدريب الأربعة لهذه الأوزان يساوي الواحد .

مثال 19:

استخدام أدالين لتنفيذ التابع **AND**، بمداخل ومخارج ثنائية القطبية، في هـذه الحالـة ستكون قيمة الأوزان التي تعطي خطأً أصغرياً مساوية $w_1 = 1/2$، $w_2 = 1/2$، $w_0 = -1/2$، وستكون معادلة خط الفصل $1/2x_1 + 1/2x_2 - 1/2 = 0$ وهو نفس خط فصل شبكة البيرسبترون.

مثال 20:

استخدام أدالين لتنفيذ التابع **OR**، بمداخل ومخارج ثنائية القطبية، ستكون قيمة الأوزان التي تعطي خطأً أصغرياً هي: $w_0 = 1/2$، $w_1 = 1/2$، $w_2 = 1/2$، وخط الفصل الموافق $1/2x_1 + 1/2x_2 + 1/2 = 0$.

مثال21:

استخدام أدالين لتنفيذ التابع **AND NOT**، بمداخل ومخارج ثنائية القطبية، ستكون قيمة الأوزان المعطية خطأً أصغرياً كالتالي: $w_0 = -1/2$، $w_1 = 1/2$، $w_2 = -1/2$، وخط الفصل الموافق $1/2x_1 - 1/2x_2 - 1/2 = 0$.

3.6.4 تطبيقات أدالين في حقل الاتصالات

استُعمل عنصر المعالجة البسيط هذا استعمالاً واسعاً في التطبيقات الهندسية مثل الترشيح المتكيف للضجيج وتضعيف الصدى وتمييز النموذج والتنبؤ.

مثال 22:

استخدام أدالين كمرشح متكيف للضجيج، في هذا التطبيق تكون إشارة الدخل عادة تشابهية (مستمرة)، لذا تُجرَى عملية أخذ عينات عليها باستخدام مبدل تشابهي رقمي **(analog-digital converter)**. تطبق المداخل x_k على مجموعة من وحدات التأخير المربوطة فيما بينها على التسلسل كما هو موضح في (الشكل 35.4)، حيث كل وحدة تقوم بتأخير

الإشـــارة خطـــوة زمنية واحدة، (Widrow & Winter) عام 1988[32].

الوحدة أدالين هي جامعٌ خطي متكيف، يوضع في المخطط بين وحدات التأخير والخرج للمرشح ليقوم بعملية جمع خطي للعينات الحالية والسابقة 0 تُعدَّل الأوزان للتحكم بالاستجابة النبضية للإشارة من الدخل إلى الخرج، والتحكم بالاستجابة النبضية يكافئ التحكم بالاستجابة الترددية. لهذا، تعدل الأوزان لتعطي إشارة الخرج أفضل قيمة للمربعات الصغرى عبر الزمن للاستجابة المرغوب بها في إشارة الدخل.

تستعمل مرشحات متكيفة من هذا النوع لإنجاز موازنة (Equalization) للقناة لتعويض تشويه النبضة الرقمية المنبعث من عناصر دارة الاتصالات الرقمية ذات مميزات الاستجابة الترددية غير الخطية. في هذه الحالة، يوضع المرشح في قناة الاتصالات ليستقبل ويوازن النبضات x_k المشوهة. يمرر خرج الموازن من خلال مكمم (quantizer) لإعادة النبضة المشوهة إلى شكلها الأصلي (بدون ضجيج).

تستطيع القناة الهاتفية النموذجية بدون موازن استرداد حوالي 90% من المعطيات المرسلة الخالية من الضجيج. وعندما يصبح المرشح المتكيف مدرباً على إلغاء الضجيج، فإن معدل الخطأ ينخفض إلى حوالي 10^{-6} (خطأ واحد لكل مليون نبضة مرسلة). يوضح (الشكل 4-36) نظام حذف للضجيج متكيف.

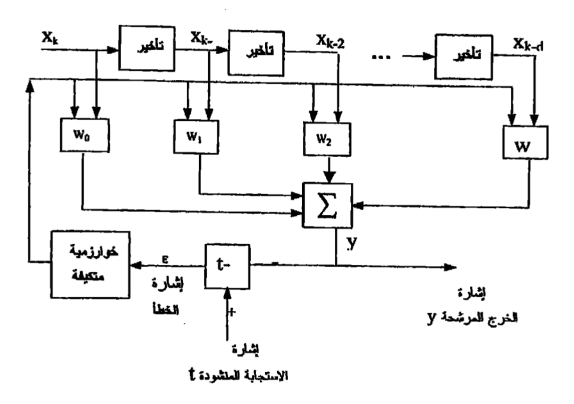
في هذا النظام إشارة الدخل الأولية تتألف من الإشارة المرغوب بها x(n) المشوهة بالضجيج $w_1(n)$ وضجيج التداخل $w_2(n)$ و$w_3(n)$.

وهكذا ستكون إشارة الخطأ كما يلي:

$$\begin{aligned} e(n) &= y(n) - w_2(n) \\ &= y(n) \sum_{k=0}^{M-1} h(k) v(n-k) \end{aligned} \qquad (51.4)$$

يستعمل هذا الخطأ لتعديل ثوابت المرشح المتكيف (Adaptive FIR) والذي هو عبارة عن شبكة أدالين. يستعمل معيار المربعات الصغرى لتحديد ثوابت المرشح. وستكون نتيجة الأمثلية مجموعةً من المعادلات الخطية التالية:

$$\sum_{k=0}^{M-1} h(k) r_{vv}(l-k) = r_{yv}(l) \quad , \quad l = 0, 1, 2, \dots, M-1 \qquad (52.4)$$



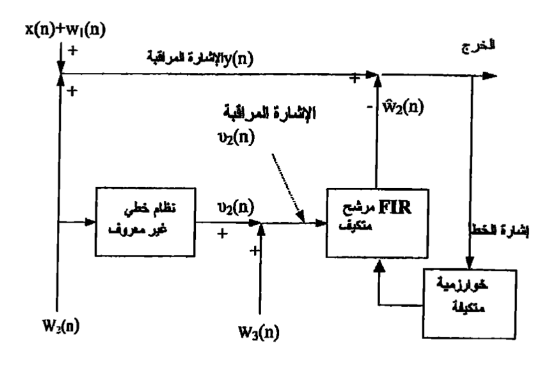
الشكل 35.4 شبكة أدالين كمرشح متكيف

حيث $r_{vv}(l)$ هو تابع الارتباط الذاتـــي للمتتالية $\upsilon(n)$ و $r_{yv}(l)$ هو تابع الارتباط المتبادل للمتتاليتين $y(n)$ و $\upsilon(n)$.

استعملت شبكة أدالين أيضاً لإلغاء الضجيج لأن الخرج الفعلي هو خرج الهدف المنشود ولأن المتوسط عبر المخارج الأساسية صحيح بنسبة 90% من الزمن.

عندما تصبح الأوزان متقاربة (تعين مجموعة الأوزان التـــي تعطي استجابة صحيحة)، ينخفض معدل الخطأ، وهذا ما يسمح بزيادة معدل بث المعطيات عبر القناة.

في الحقيقة عندما يُستعمل الموازن مع المودم تزداد مقادير البث المسموح بها أربعة أضعاف عن المقادير في حالة عدم استعمال موازن، وذلك من أجل نفس مستوى الإنجاز (سنرسم المخطط لاحقاً).



الشكل 36.4: نظام إلغاء ضجيج متكيف

إن التكلفة المالية التـــي ندفعها لتزويد القناة بموازن صغيرة جداً إذا ما قورنت بما ندفعه من تكلفة مالية لزيادة عرض قناة البث.

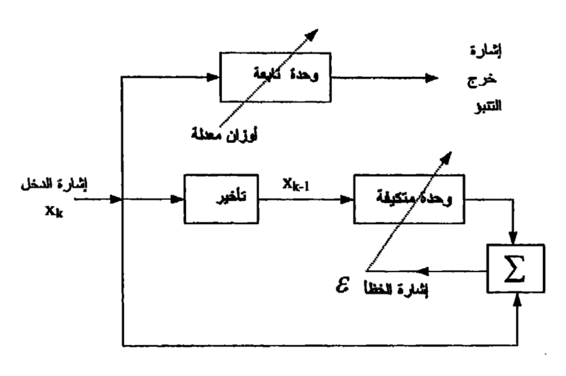
مثال 23:

استخدام أدالين متنبئاً متكيفاً، تطبيق هام آخر للترشيح المتكيف هو التنبؤ بالإشارة. في هذه الحالة تستعمل العينات الحالية للسلسلة الزمنية التشابهية الارتباطية كخرج منشود، وتصبح عينات الزمن المتأخرة نموذجاً للدخل. في هذه الوحدة الموضحة في (الشـــكل 37.4) والمقترحة مـــن قـــبل Widrow & winter، عام 1988[32]، ستكون عينة الإشارة المتأخرة دخل الشبكة.

مثال 24:

استخدام أدالين لحذف الصدى، مثال آخر للشبكة أدالين في ربوع حقل الاتصالات، هو استخدام هذه الشبكة لإزالة الصدى (Echo cancellation) على نحو متكيف في دارات الاتصالات ذات المسافات الطويلة. تُجرى الاتصالات ثنائية الاتجاه اللحظية وفقاً لطاقة الربط بين المرسل والمستقبل عند نهايتـــي الدارة طويلة المسافة. يسبب أحياناً هذا الربط انعكاساً

للطاقة من المستقبل إلى المرسل (سببه على الأغلب عدم تلاؤم ممانعات دخل وخرج الأدوات الإلكترونية في دارة الربط).

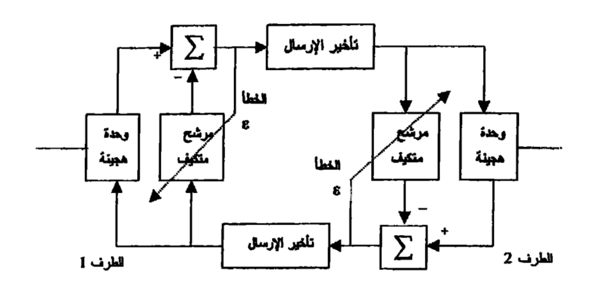


الشكل 37.4: متنبئ متكيف

يسمى الصدى عند المستقبل الآتـــي من دارة الاستقبال بصدى النهاية القريبة (near-end echo)، ويسمى الصدى عند المستقبل الآتـــي من دارة الإرسال بصدى النهاية البعيدة (far-end echo). يوجد كلا نوعي الصدى في المعطيات المرسلة عبر الخط الهاتفي، وعلينا تصميم مزيل الصدى لحذف أو تضعيف هذا الصدى إلى أقل ما يمكن.

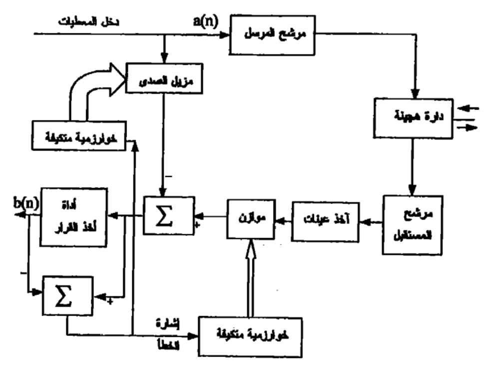
حلت مشكلة الخط الهاتفي ثنائي السلك باستعمال مزيل للصدى متكيف (أدالين) عند طرفي قناة الاتصال، كما هو موضح في (الشكل 38.4)، (Widrow & winter عام 1988[32]).

الوحدات لتعمل كدارة هاتف محلي. ولكن تبعاً للممانعات المختلفة بين عناصر الدارة ككل، تسمح الوحدات الهجينة نفسها لمقدار من الإشارة بالعودة إلى طرف الإرسال وهذا يؤدي إلى عودة إشارة الصدى. تعوض المرشحات المتكيفة التصميم غير المتلائم بين المحولات



الشكل 38.4: دارة اتصال للمسافات البعيدة مع إزالة للصدى

الهجينة والخواص الكهربائية المختلفة للدارة المحلية. تقوم هذه المرشحات بإلغاء أي إشارة منتشرة في الاتجاه المعاكس من دارة المرسل إلى دارة المستقبل؛ ومن ثم تضعف الصدى تضعيفاً كبيراً.



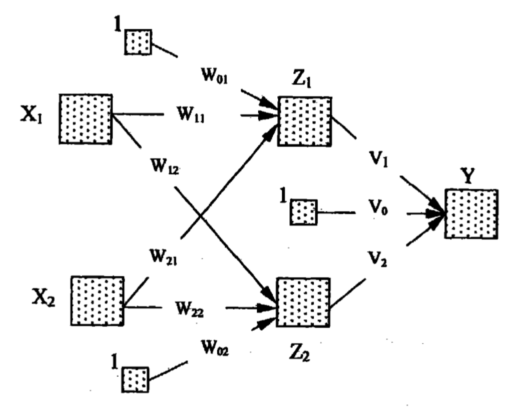
الشكل 39.4: مودم مع موازن ومزيل للصدى متكيف

لقد انتشر استعمال وحدات المرشحات المتكيفة (وحدات أدالين) في دارات الاتصال وذلك بعد انتشار الخطوط العابرة للقارات. وأخيراً سنبين في (الشكل 39.4) مخطط مودم

يستخدم مزيل الصدى مع الموازن لتعويض تشويه النبضة الرقمية في دارات الاتصالات الرقمية المشروحة سابقاً [251].

7.4 شبكة MADALINE

ذكرنا سابقاً، أن شبكة MADALINE تتألف من تركيب عدة شبكات خطية متكيفة (عدة وحدات أدالين) مرتبة في طبقات عديدة. يبين (الشكل 40.4) بنية شبكة مادالين (MADALINE) بسيطة مؤلفة من طبقة مخفية واحدة تحتوي على وحدتي أدالين هما: Z_2, Z_1.



الشكل 40.4: شبكة **MADALINE** بطبقة مخفية واحدة

مدخل كل وحدة أدالين له نموذج دخل وانحياز ثابت $(x_1, x_2, 1)$. يحسب المجموع الخطي المثقل للمداخل net بواسطة كل أدالين وتمرر النتيجة إلى وحدة بعتبة ثنائية القطبية. بعدئذ يمرر المخرجان (z_1, z_2) إلى وحدة منطقية مفردة AND (العصبون Y، ثنائي القطبية) التـــي ستعطي ‎-1 عندما يكون لوحدتـــي أدالين الدخل نفسه، وستعطي ‎+1 عندما يكون المدخلان مختلفين (وهذا هو التابع XOR، سنعطي بعض الأمثلة فيما بعد).

يعطي استخدام الوحدات المنخفية Z2 و Z1 للشبكة مقدرات حسابية تفتقر إليها الشبكات

وحيدة الطبقة، ولكن أيضاً تتعقد خوارزمية التعليم لمثل هذه الشبكات. تعتبر طريقة تشكيل عدة طبقات من وحدات أدالين للحصول على مادالين متعدد الطبقات إحدى الطرق المتبعة للتغلب على المشاكل المرافقة للتطبيقات الحسابية التـي تتطلب إمكانيات فصل غير خطية. فمثلاً باختيار مناسب للأوزان فإن شبكة مادالين البسيطة الموضحة في (الشكل 40.4) تصبح قادرة على حل مشكلة XOR التـي واجهناها في مهام تصنيف النماذج باستخدام شبكة أدالين وأخواتها وحيدة الطبقة.

1.7.4 خوارزميات تعليم مادالين

قاعدة تكيف مادالين MR (MADALINE adaptation Rule)هي لون من ألوان التعليم بمعلم. وهي مبنية على مبدأ تعديل الأوزان لنموذج التدريب الحالي مع إجراء يبدو وكأنه تمزيق صغير للتمثيلات المُتعلّمة لنماذج التدريب السابقة. أطلق Widrow على هذا المبدأ اسم مبدأ الاضطراب الأصغري (minimal disturbance principal).

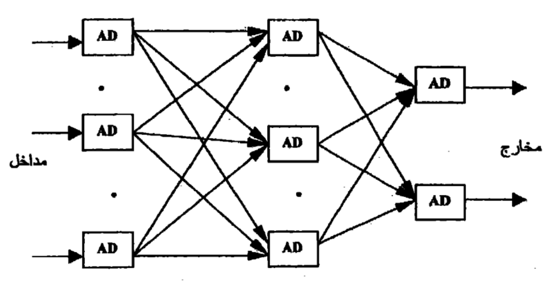
يمكن التعبير عن هذا المبدأ كما يلي: ليكن لدينا نموذج تدريب معطى، يبدأ الإجراء باختيار عصبون في أول طبقة مخفية له استجابة خرج تشابهية **قريبة من الصفر**، الآن تُغيَّر أوزان هذا العصبون **لتعكس** استجابة الخرج الثنائية. ينتشر هذا التغير إلى عصبونات الخرج حيث يجري فحص سريع لمعرفة: هل **يخفّض هذا التغير القيمة الكلية لأخطاء الخرج**؟ إذا كان الأمر كذلك فالتغير يكون مقبولاً، وإلا أُعيد تخزين قيمة الأوزان السابقة واختير عصبون آخر في نفس الطبقة الأولى ذو استجابة الخرج التالية **الأقرب إلى الصفر** لتعديل أوزانه. تستمر هذه العملية حتى يتم اختيار كل العصبونات في الطبقة الأولى للملاءمة.

بعدئذ نختار أزواجاً من العصبونات في الطبقة الأولى للملاءمة، باستعمال نفس معيار اختيار استجابة الصفر الأقرب (nearest zero). يُفحص الزوج الملائم وأخطاء الخرج المرافقة لمعرفة: هل خُفّضت قيمة الخطأ الكلي؟ إذا كان كذلك فالتغيرات تكون مقبولة، وإلا أُعيدت الأوزان إلى قيمتها الأصلية واختير زوج آخر. بعد ملاءمة كل العصبونات في الطبقة الأولى زوجاً زوجاً، يمكن اختيار ثلاث عصبونات للملاءمة إذا كان ذلك لازماً (عادة يكتفى بعملية الاختبار بالأزواج لتحقيق المراد)، نقوم باختبار الطبقة الثانية للملاءمة وتتبع نفس

الإجراءات كما في الطبقة الأولى. تتكرر هذه العملية طبقة تلو الأخرى حتى نصل إلى طبقة الخرج. يبين (الشكل 41.4) شبكة مادالين متعددة الطبقات المخفية.

تُدرَّب العصبونات في طبقة الخرج، على نحو مختلف عنه في الطبقات المخفية، وذلك باستعمال قاعدة تعليم دلتا أو طريقة تعليم بمعلم أخرى لأن الأخطاء معروفة لكل وحدة أدالين على حدة في نفس الطبقة.

بعد تدريب كل وحدات أدالين في الشبكة لنموذج التدريب الأول، يختار نموذج التدريب الثاني للدخل وتستمر العملية هكذا حتى تطبق كل نماذج دخل التدريب. باختيار وحدات أدالين بمجاميع تشابهية أقرب إلى الصفر، يجري أولاً اختيار الوحدة التـــي تعطي استجابة معاكسة، ومن ثم تلك التـــي لها التأثير الأعظمي على الخرج وفقاً لتغير أصغري لقيمة الوزن (تحقيق مبدأ الاضطراب الأصغري من أجل التأثير الأعظمي).



الشكل 41.4: مادالين متعدد الطبقات من أدالين

لقـــد عُدِّلـــت الخوارزمية MRI وأعيد تعريفها عـــبر الزمـــن لكل نســـخة جديـــدة (MRIII ,MRII ,MRI). الخوارزمية المشروحة آنفاً لـــ MRII، والخوارزمية MRIII تقريباً هي نفس MRII باستثناء أنها تطبق على شبكات متعددة الطبقات بتوابع تفعيل Sigmoid عوضاً عن توابع تفعيل منطقية العتبة (Andes عام 1991[33]).

سنشرح خوارزميتـــي تعليم شبكة مادالين بطبقة مخفية واحدة فقط، حيث تقوم

الخوارزمية الأساسية الأولى بتعديل **أوزان الطبقة المخفية فقط** w_{ij}، أما أوزان **وحدة الخرج** υ_i **فتبقى ثابتة**، وتسمى هذه الخوارزمية MRI (الشكل الأصلي لتدريب MADALINE، المقترحة من قبل Widrow & Hoff 1960[29])، أما الخوارزمية المعدلة MRII عن الخوارزمية الأساسية (والمقترحة من قبل Widrow & Winter & Baxter 1987[34]) **فإنها تقوم بتعديل جميع الأوزان في الشبكة.**

1.1.7.4 خوارزمية MRI لتعليم مادالين بطبقة مخفية واحدة

الشبكة المدربة موضحة في الشكل الســـابق (40.4). وســـيكون تابع التفعيل للوحدات Z_1, Z_2, Y هو:

$$f(net)=\begin{cases}+1 & , \text{ if } net\geq 0\\ -1 & , \text{ if } net<0\end{cases} \qquad (53.4)$$

1. تعطى الأوزان قيماً صغيرة كيفية، وعادة تختار الأوزان $\upsilon_1, \upsilon_2, \upsilon_0$ مساوية لأوزان أدالين لتحقيق مربع خطأ أعظمي (ولا تُعدَّل هذه الأوزان خلال الخوارزمية، كما ذكر في النص آنفاً). α معدل التعليم، كما في خوارزمية تدريب أدالين، يعطى قيمة صغيرة.
2. مادام شرط التوقف عن التدريب غير محقق، كرر الخطوات من 3 إلى 9.
3. لكل زوج تدريب (s, t)، كرر الخطوات من 4 إلى 8.
4. ضع التفعيلات لوحدات الدخل $x_i = s_i$.
5. احسب الدخل net لكل وحدة أدالين ADALINE مخفية:

$$\begin{aligned} net\, z_1 &= w_{01} + x_1 w_{11} + x_2 w_{21} \\ net\, z_2 &= w_{02} + x_1 w_{12} + x_2 w_{22} \end{aligned} \qquad (54.4)$$

6. حدد خرج كل وحدة أدالين ADALINE مخفية :

$$\begin{aligned} z_1 &= f(net\, z_1) \\ z_2 &= f(net\, z_2) \end{aligned} \qquad (55.4)$$

7. حدد خرج شبكة مادالين MADALINE

(لاحظ أن قيم أوزان وحدة الخرج υ_0, υ_2, υ_1 تبقى ثابتة خلال الخوارزمية):

$$net\,y = v_0 + z_1 v_1 + z_2 v_2$$
$$y = f(net\,y) \quad (56.4)$$

8. احسب الخطأ، وعدِّل الأوزان كما يلي:

إذا كان $t = y$، أي لا يوجد أخطاء فلا تعديل للأوزان وإلا إذا كان $t = 1$ غيِّر أوزان الوحدة Z_j التـــي لها الدخل الأقرب إلى الصفر:

$$w_{oj}^{new} = w_{oj}^{old} + \alpha(1 - net\,z_j)$$
$$w_{ij}^{new} = w_{ij}^{old} + \alpha(1 - net\,z_j)\,x_i \quad (57.4)$$

إذا كان $t = -1$ غيِّر أوزان كل الوحدات Z_k التـــي لها دخل موجب:

$$w_{ok}^{new} = w_{ok}^{old} + \alpha(-1 - net\,z_k)$$
$$w_{ik}^{new} = w_{ik}^{old} + \alpha(-1 - net\,z_j)\,x_i \quad (58.4)$$

9. شرط التوقف:

إذا توقف تغير الأوزان (أو وصل إلى مستوى مقبول)، أو أُنجز العدد الأعظمي المخصص لمرات تكرار تعديل الأوزان (الخطوة 2)، عندئذ توقف، وإلا استمر في تنفيذ الخوارزمية.

2.1.7.4 خوارزمية MRII لتعليم مادالين

يجري في هذه الخوارزمية تعديل لجميع أوزان الشبكة **w** و**v** وتحقيق مبدأ الاضطراب الأصغري. إن خطوات هذه الخوارزمية مشابهة تماماً للخوارزمية السابقة، لذا لن نعيد كتابة الصيغ، وسنقوم بشرح مثال توضيحي فيما بعد. ستكون خطوات الخوارزمية كما يلي:

1. أعطِ لجميع أوزان الشبكة **w** و**v** قيماً أولية، وحدِّد قيمة α معدل التعليم.
2. مادام شرط التوقف غير محقق كرِّر الخطوات من 3 إلى 9.
3. كرِّر الخطوات من 4 إلى 8 لكل زوج تدريب ثنائي القطبية **(s ,t)**.

7.4 احسب خرج الشبكة net كما في خوارزمية MRI السابقة.

8. عيِّن الأخطاء وحدِّث كافة الأوزان **w** و**v** إذا كان ذلك ضرورياً:

إذا كان $t \neq y$ كرر الخطوات 8 أ- ب لكل وحدة مخفية لها دخل شبكة net قريب قرباً كافياً من الصفر (افترض أن دخل الشبكة سيكون ضمن المجال [0.25, 0.25–]. ابدأ بالوحدة

التـــي لها دخل شبكة أقرب إلى الصفر،من أجل هذا الأقرب، ...الخ

8. أ غيِّر خرج الوحدة j من +1 إلى -1 أو بالعكس.

8. ب أعِدْ حساب استجابة الشبكة.

إذا انخفضت قيمة الخطأ عدِّل الأوزان على الوحدة (استعمل قيمة الخرج الجديدة بعد تعيينها كخرج منشود وطبِّق قاعدة دلتا).

9. اختبر شرط التوقف:

إذا توقف تغير الأوزان (أو وصل إلى المستوى المقبول)، أو أُنجز العدد الأعظمي المخصص لمرات التكرار (الخطوة 2)، فتوقّف وإلا استمر في تنفيذ الخوارزمية.

2.7.4 تطبيقات مادالين في تصنيف النماذج

مثال 25:

استخدام مادالين لتنفيذ التابع المنطقي **XOR**، بعد أن استعرضنا خوارزميات تدريب مادالين سنقوم بتدريب هده الشبكة على حل مشكلة XOR التـــي استعصت على شبكة البيرسبترون PERCEPTRONE وأخواتها وحيدة الطبقة. سنستخدم خوارزمية MRI لعملية التدريب هذه، وسنعرض تحديث أو تغيير الأوزان بالتفصيل لأول تحديث فقط. تعطى نماذج التدريب بجدول الحقيقة التالي:

x_1	x_2	T
1	1	−1
1	−1	1
−1	1	1
−1	−1	−1

الشبكة موضحة في (الشكل 40.4). سنفرض القيم الأولية للأوزان w_{ij} كما يلي:

الأوزان إلى الوحدة Z_1: $w_{11} = 0.05 \quad ,w_{21} = 0.2 \quad ,w_{01} = 0.3$

الأوزان إلى الوحدة Z_2: $w_{22} = 0.2 \quad ,w_{12} = 0.1 \quad ,w_{02} = 0.15$

سنعتمد أوزان الوحدة Y لتكون مساوية للأوزان في المثال 19 لشبكة أدالين وذلك للحصول على أقل مربع خطأ ممكن.

أوزان الوحدة Y: $v_1 = 0.5 \quad , v_2 = 0.5 \quad , v_0 = 0.5$

و قيمة $\alpha = 0.5$، مع تابع تفعيل وحدات أدالين الثلاثة Z_1, Z_2, Y:

$$f(net) = \begin{cases} +1 & , \quad if net \geq 0 \\ -1 & , \quad if\ net < 0 \end{cases}$$

في حالة أول زوج تدريب (1, 1): $x_1 = 1 \quad , x_2 = 1$

$$net\, z_1 = 0.3 + 0.05 + 0.2 = 0.55$$
$$net\, z_2 = 0.15 + 0.1 + 0.2 = 0.45$$
$$z_1 = 1$$
$$z_2 = 1$$
$$net\, y = 0.5 + 0.5 + 0.5 = 1.5$$
$$y = 1$$
$$t - y = -2 \neq 0$$

هذا يعنـــي حدوث خطأ، وباعتبار أن $t = -1$ وكلا الوحدات Z لها دخل شبكة موجب، سيكون تحديث الأوزان على Z_1 كما يلي:

$$w_{01}^{new} = 0.3 + (0.5)(-1.55) = -0.475$$
$$w_{11}^{new} = 0.05 + (0.5)(-1.55) = -0.725$$
$$w_{21}^{new} = 0.2 + (0.5)(-1.55) = -0.575$$

وسيكون تحديث الأوزان على Z_2 كما يلي:

$$w_{02}^{new} = 0.15 + (0.5)(-1.45) = -0.575$$
$$w_{12}^{new} = 0.1 + (0.5)(-1.45) = -0.625$$
$$w_{22}^{new} = 0.2 + (0.5)(-1.45) = -0.525$$

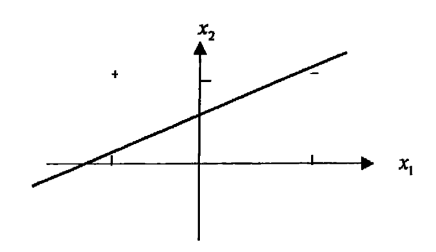
بعد أربعة أدوار من التدريب، ستكون قيمة الأوزان النهائية مساوية لـــ:

$$w_{01}^{new} = -0.99 \quad , \quad w_{02}^{new} = -1.09$$
$$w_{11}^{new} = -0.73 \quad , \quad w_{22}^{new} = -1.33$$
$$w_{21}^{new} = 1.53 \quad , \quad w_{12}^{new} = 1.27$$

في حالة هذه الأوزان النهائية يمكن تمثيل منطقة الاستجابة الموجبة للشبكة. إن منطقة الاستجابة لشبكة مادالين مدربة هي اجتماع مناطق الاستجابة الموجبة لوحدتّي أدالين Z_2, Z_1

المخفية. مثلاً، ستكون معادلة خط الفصل للوحدة المخفية Z_1 الموضح في (الشكل 42.4) كما يلي:

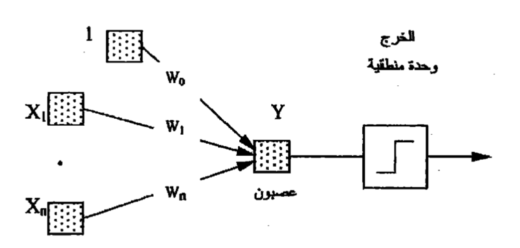
$$x_2 = 0.48x_1 + 0.65$$



الشكل 42.4: منطقة الاستجابة الموجبة للوحدة Z_1

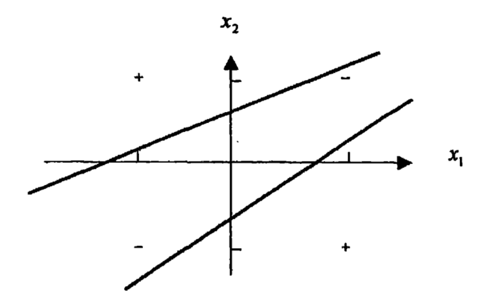
وستكون معادلة خط الفصل للوحدة المخفية Z_2 الموضح في (الشكل 43.4) كما يلي:

$$x_2 = 0.96x_1 - 0.82$$



الشكل 43.4: منطقة الاستجابة الموجبة للوحدة Z_2

ويبين (الشكل 44.4) منطقة الاستجابة الموجبة لشبكة مادالين للتابع XOR.



الشكل **44.4**: منطقة استجابة التابع المنطقي **XOR**

3.7.4 تطبيقات مادالين في تعرف الأشكال

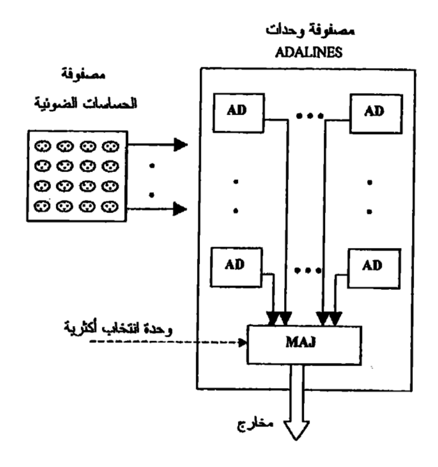
إن شبكات مادالين قادرة أيضاً على التعلم لحساب أي تابع بسلوك حسن؛ وقد أنشئت لصنف من الشبكات العصبونية الصنعية المتعددة الطبقات بتغذية أمامية MLFF (MultiLayer FeedForward). سنقوم الآن بعرض لتطبيق مرتبط بالرؤية؛ وهو عبارة عن نوع من تمييز الأشكال غير المتغيرة، ستقوم الشبكة بتعرف الأشكال مستقلة بعضها عن بعض بالمكان والاتجاه.

بالطبع تتطلب هذه المهمة الصعبة من الشبكة أولاً أن تتعلم تمييز الأشكال المختلفة فيما بينها، مثلاً بعض الأماكن في الصورة المرئية، ومن ثم أن تكون قادرة على تعرف نفس الشكل عندما يظهر في مكان أو اتجاه آخر في الصورة (وكأن الشبكة يراد منها مراقبة حركة أحجار الشطرنج مثلاً؛ أولاً هناك مهمة تعرّف شكل الأحجار ومن ثم معرفة الحجر عندما يظهر في مكان ما من الرقعة)، وهذا ما نعني بعبارة عدم التغير، أي إن شكل الشيء ثابت والمتغير هو المكان.

مثل هذه المهمة تتطلب شبكة تحتوي على عدد من المصفوفات المربعة (الألواح)، حيث عناصر كل مصفوفة هي وحدات أدالين بوصلات داخلية خاصة. البنية الأساسية كما

وصفها Widrow وWinter عام 1988 لمصفوفة واحدة فقط موضحة في (الشكل 45.4) المقترح من قبل Widrow & Winter عام 1988[32].

تتألف طبقة الدخل من مصفوفة مربعة من الخلايا الحساسة للضوء، ويعتمد حجم هذه المصفوفة على التحليل اللازم للصورة. إن وحدة أدالين في الزاوية العليا اليسرى من المصفوفة الأولى (سنرمز لها AD_{11}) سيكون لها مصفوفة أوزان مربعة W_1 على وصلاتها الداخلية. ويمكن أن تكون هذه الأوزان ذات قيم كيفية.



الشكل 45.4: مصفوفة العصبونات لإزاحة لاتغيرية يمين ـ يسار، وأعلى ـ أسفل

إن مصفوفة الأوزان لوحدة أدالين (AD_{12}) إلى اليمين تماماً من وحدة الزاوية اليسرى العليا (AD_{11}) لها نفس قيم مصفوفة أوزان AD_{11}، ولكنها حولت كمجموعة لنأخذ بعين الاعتبار إزاحة عنصر صورة (pixel) واحد إلى اليمين. سنسمي هذا التحويل $T_{R1}(W_1)$، وهكذا سنحول المصفوفة W_1 إلى اليمين لكل وحدة على السطر الأول وحدة تلو وحدة؛ فالوحدة AD_{12} سيكون لها مصفوفة وزن $T_{R2}(W_1)$، والوحدة AD_{13} لها مصفوفة وزن

$T_{R3}(W_1)$ وهكذا حتى نهاية السطر الأول. يجب ملاحظة أن التحويلات تلتف دائرياً؛ أي إن آخر وحدة في أقصى يمين السطر الأول تزاح بتحويل واحد عن الوحدة AD_{11} في أقصى يسار السطر الأول.

أما الوحدة AD_{21} التـــي تأتـــي في السطر الثاني أسفل AD_{11} مباشرة فسيكون لها أوزان محولة لتمثيل إزاحة سفلية بمقدار عنصر صورة واحد باستعمال التحويل $T_{D1}(W_1)$. والوحدة AD_{31} التـــي تأتـــي في السطر الثالث أسفل AD_{21} مباشرة لها مصفوفة أوزان $T_{D2}(W_1)$ وهكـــذا. يمكننا الآن التعبير عن أوزان وحـــدات اللـــوح الأول الـــذي يحوي $(n + 1) \times (m + 1)$ وحدة بالشكل المصفوفي التالي:

$$(W1)\ T_{R1}(W_1)\ T_{R2}(W_1)\ ....\ T_{Rn}(W_1)$$

$$T_{D1}(W_1)\ T_{R1}\ T_{D1}(W_1)\ T_{R2}\ T_{D1}(W_1).....\ T_{Rn}T_{D1}(W_1)$$

$$T_{D2}(W_1)\ T_{R1}\ T_{D2}(W_1)\ T_{R2}\ T_{D2}(W_1).....\ T_{Rn}\ T_{D2}(W_1)$$

$$............\quad ....\quad ....\ ....$$

$$T_{Dm}(W_1)\ T_{R1}\ T_{Dm}(W_1)\ T_{R2}\ T_{Dm}(W_1)\ .....\ T_{Rn}\ T_{Dm}(W_1)$$

إن خرج جميع الوحدات في المصفوفة (اللوح) الواحدة موصل إلى وحدة عتبة ذات انتخاب أكثرية تدعى MAJ (MAJority)، حيث يكون خرج هذه الوحدة موجباً (للّوح الواحد) عندما يستجيب نصف الوحدات على الأقل يستجيب إيجابياً لنموذج الدخل. ولما كانت الوحدة MAJ تعالج كل الوحدات في اللوح الواحد بالتساوي، ومن ثم فإن الخرج سيبقى غير متغير لنموذج الدخل، وعلى نحو مستقل عن مكانه. عندما ينتقل الشيء يميناً ويساراً أو للأعلى والأسفل، فإن أدوار الوحدات تتغير ولكن خرج اللوح يبقى نفسه.

أما بالنسبة للوح الثاني فنستعمل أيضاً التحويلات السابقة نفسها ولكن لقيم مختلفة لمصفوفة الأوزان W_2 عن قيم مصفوفة اللوح الأول. يستجيب هذا اللوح عموماً على نحو مختلف عن اللوح الأول، لكنه يبقى غير متغير للتنقلات في نموذج الدخل. وهناك ألواح أخرى نظمت بنفس الطريقة كما في اللوحين السابقين، لكن بمصفوفات وزن أولية مختلفة.

إن مخارج كل وحدات MAJ الخاصة بكل لوح موصلة إلى شبكة مادالين متكيفة، تعمل

هذه الشبكة على فك التجميع (فارزة) (descrambler)، يمكن أن تدرب لتعطي استجابة ملائمة لنماذج مختلفة باستعمال واحد من خوارزميات تدريب MR الموصوفة سابقاً.

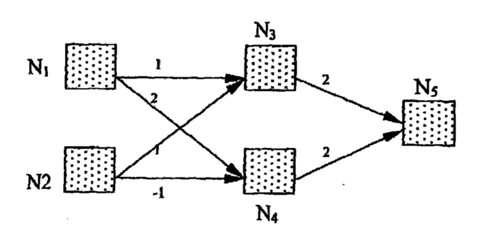
سنختتم هذا الفصل بالإشارة إلى أن الشبكات العصبونية الصنعية مادالين التشابهية تطورت تطوراً كبيراً وصممت على شرائح صُلبة VLSI، حيث تستعمل كثيراً في التطبيقات العسكرية المتنوعة (Widrow وزملاؤه عام 1994[35]).

4.7.4 تمارين

1.4 لتكن لدينا شبكة عصبونية مؤلفة من عصبونات McCulloch-Pitts المبينة في (الشكل 46.4). لكل عصبون عتبة بقيمة 2.

1. عرِّف استجابة العصبون N_5 عند اللحظة t بدلالة تفعيلات عصبونات الدخل N_1 وN_2 عند الزمن المناسب.

2. استنتج تفعيل كل عصبون ناتج من إشارة دخل $N_2 = 0$, $N_1 = 1$ عند اللحظة $t = 0$.



الشكل 46.4: شبكة عصبونية للتمرين 4-1

2.4 هناك على الأقل طريقتان لل تعبير عن التابع XOR بدلالة التوابع المنطقية البسيطة يمكن تمثيلها بعصبونات McCulooch-Pitts. إحدى هذه الطرق شُرحت في الفصل الرابع. أوجد التمثيل الآخر والشبكة المقابلة. وقارن بين الشبكتين المعطاة والمستنتجة.

3.4 في نموذج McCulloch-Pitts للتحسس بالحرارة والبرودة، منبهات البرودة التي طبقت عند اللحظات $t-2$ و$t-3$ شعر بها كبرودة في اللحظة t. هل تستطيع تعديل الشبكة لكي

نطبق منبهات البرودة ثلاث خطوات زمنية قبل أن نشعر بالبرودة؟

4.4 صمم شبكة McCulloch-Pitts لنمذجة متحسس بنماذج موسيقية بسيطة. استعمل ثلاث وحدات دخل لتوافق النغمات الثلاث "do"، "re"، "mi". افترِض أن نغمة واحدة قدمت للشبكة عند أية لحظة. استعمل وحدتين للخرج لتوافق التحسس بـــ "مقطع أعلى- السلم" و"مقطع أسفل - السلم"، وخاصة.

1. نموذج المداخل "do" في اللحظة $t = 1$، و"re" في اللحظة $t = 2$، و"mi" في اللحظة $t = 3$ سينتزع استجابة موجبة من الوحدة "مقطع أعلى- السلم".

2. نموذج المداخل " mi" في اللحظة $t = 1$، و"re" في اللحظة $t = 2$، و"do" في اللحظة $t = 3$ سينتزع استجابة موجبة من الوحدة "مقطع أسفل- السلم".

3. أي نموذج من المداخل لن يولد أية استجابة قد ترغب في توسيع هذا المثال، سامحاً لأكثر من وحدة دخل واحدة "on" عند أي لحظة زمنية، صمم وحدات الخرج لكشف النغمات.

5.4 طبق قاعدة تعليم Hebb على نماذج التدريب التـــي تعرف التابع XOR.

6.4 هناك 16 تابعاً منطقياً (بمدخلين ومخرج وحيد)، 14 من هذه التوابع مفصولة خطياً. أثبت أن قاعدة Hebb يمكن أن توجد الأوزان لجميع المسائل المستخدمة فيها هذه التوابع، استخدم التمثيل ثنائي القطبية مع الانحياز.

7.4 باستعمال قاعدة Hebb:

1. أوجد الأوزان اللازمة لإنجاز التصانيف التالية: الأشعة $(1, 1, 1, 1)$ و$(-1, 1, -1, -1)$ هي أعضاء من الصف (لذا قيمة الخرج ستكون 1)، والأشعة $(1, 1, 1, -1)$ و$(1, -1, -1, 1)$ ليست أعضاء من الصف (قيمة الخرج ستكون -1).

2. باستعمال كل أشعة التدريب **x** كدخل، اختبر استجابة الشبكة.

8.4 أحياناً تستعمل قاعدة Hebb بواسطة عكس نماذج التدريب الثنائية (مداخل ومخارج) إلى الشكل الثنائي القطبية لإيجاد مصفوفة الأوزان.

1. طبق هذا الإجراء لإيجاد الأوزان اللازمة لتحقيق التصانيف التالية:

$$s(1) = (1, 0, 1)\ t(1) = 1$$

$s(2) = (1, 1, 0)\ t(2) = 0$

2. باستعمال تابع الخطوة الثنائي (بعتبة مساوية للصفر) كتابع تفعيل وحدة الخرج، اختبر استجابة شبكتك على كل نماذج التدريب الثنائي.
3. باستعمال تابع الخطوة الثنائي القطبية (بعتبة صفرية) كتابع تفعيل وحدة الخرج، اعكس نماذج التدريب إلى الشكل الثنائي القطبية واختبر استجابة الشبكة ثانية.
4. اختبر استجابة شبكتك على كل النسخ الضجيجية التالية لنماذج التدريب ثنائية القطبية:

(0, −1, 1) (0, 1, −1) (0, 0, 1) (0, 0, −1) (0, 1, 0) (0, −1, 0)

(1, 0, 1) (1, 0, −1) (1, −1, 0) (1, 0, 0) (1, 1, 0) (1, 1, 1)

أية استجابات تكون صحيحة، وأية استجابات تكون غير صحيحة، وأية استجابات غير معرفة.

9.4 باستعمال بيرسبترون نفذ التابع AND في حالة المداخل والمخارج ثنائية القطبية، ومعدل التعليم $\alpha = 1$ وقيمة العتبة والأوزان الأولية تساوي الصفر. وارسم التغيرات في خطوط الفصل.

10.4 اكتشف تأثير قيمة معدل التعليم في سرعة تقارب تعليم بيرسبترون:
1. خُذْ قيماً مختلفة لـــ α في التمرين السابق، واشرح نتائجك.
2. عدل برهان نظرية تقارب قاعدة تعليم بيرسبترون في حال قيم كيفية لـــ α.

11.4 أثبت أن استعمال الانحياز أساسي في مسألة تنفيذ التابع AND باستعمال بيرسبترون في حالة مداخل ثنائية ومخارج ثنائية القطبية. أي، أثبت أنه من المستحيل إيجاد الأوزان w_1 وw_2 التـــي تكون فيها النقاط (0, 0), (0, 1) , (1, 0), (1, 1) مصنفة تصنيفاً صحيحاً. أولاً، أثبت أن (0, 0) لن يصنف مطلقاً على نحو صحيح، وفي الواقع، لا يحدث تعليم من أجل هذه النقطة. بعدئذ، أهمل (0, 0) واختر أي زوج من (0, 1), (1, 0), (1, 1) يمكن أن يكون مصنف تصنيفاً صحيحاً. أي ابحث عن الأوزان w_1 وw_2 المحققة:

$$(1)w_1 + (1)w_2 > \theta > 0$$
$$(1)w_1 + (0)w_2 < -\theta < 0$$
$$(0)w_1 + (1)w_2 < -\theta < 0$$

12.4 أثبت أن الأوزان الأولية الصغيرة ما تزال مسموحاً بها لأي مكان من خط الفصل الأولي للبيرسبترون.

13.4 كرر مسألة استعمال البيرسبترون لتنفيذ التابع AND في حالة المداخل ثنائية والمخارج ثنائية القطبية، وأثبت أنه لا تغير في عملية التدريب إذا كانت قيمة العتبة تساوي الصفر. أثبت أن خط الفصل سيكون:

$$x_2 = -2/3 x_1 + 4/3$$

14.4 لاحظ بدقة الفرق في ما نستطيع حله باستعمال توابع التفعيل التالية:

$$1 - f(net) = \begin{cases} 1 \quad , & if \quad net \geq \theta \\ 0 \quad , & if \quad net < \theta \end{cases}$$

$$2 - f(net) = \begin{cases} 1 \quad , & if \quad net \geq \theta \\ -1 \quad , & if \quad net < \theta \end{cases}$$

$$3 - f(net) = \begin{cases} 1 \quad , & if \quad net \geq \theta \\ 0 \quad , & if \quad -\theta < net < \theta \\ -1 \quad , & \text{if} \quad \text{net} < \theta \end{cases}$$

15.4 باستعمال قاعدة تعليم بيرسبترون، أوجد الأوزان اللازمة لإنجاز التصانيف التالية: الأشعة (1, 1, 1, 1) و(1-, 1, -1, -1) هي أعضاء من الصف (قيمة الخرج تساوي 1)، والأشعة (1-, 1, 1, 1) و(1, -1, -1, 1) ليست أعضاء من الصف (الخرج يساوي 1-). استعمل معدل تعليم يساوي 1 والقيم الأولية للأوزان تساوي الصفر. استعمل كل أشعة تدريب x كدخل، اختبر استجابة الشبكة.

16.4 كرر مسألة تنفيذ التابع AND وNOT وOR باستعمال مادالين لأشعة ثنائية عوضاً عن ثنائية القطبية.

17.4 صمم شبكة متعددة الطبقات بوحدتين مخفيتين قادرة على تعلم إعطاء نموذج دخل ثنائي على نحو صحيح تماماً. الوحدة المخفية الأولى ستكون أوزانها مساوية لنموذج الدخل وعتبتها تساوي عدد مرات تكرار الواحد في نموذج الدخل. والوحدة المخفية الثانية ستصمم بحيث ستتفعل إذا كان net أقل من أو يساوي عدد مرات تكرار الواحد في النموذج المعطى. ضم خرج كلا الوحدتين المخفيتين بحيث تتفعل وحـــدة الخرج إذا

كانت كلا الوحدتين المخفيتين "on". الهدف من هذا التمرين ملاحظة أنه إذا كان p نموذج تدريب دخل، فإن 2p وحدة مخفية ستسمح للشبكة بتعلم كل نموذج تدريب بوجه صحيح تماماً.

18.4 التابع المنطقي XOR يمكن تمثيله كما يلي:

$$x_1 XOR\ x_2 \Leftrightarrow (x_1\ OR\ x_2)\ AND\ NOT(x_1\ AND\ x_2)$$

صمِّم مادالين لتنفيذ هذه الصيغة للتابع XOR، وقارن هذه الشبكة مع الشبكة المستنتجة بتطبيق قاعدة تعليم MRI المشروحة سابقاً.

19.4 باستعمال قاعدة تعليم Delta، أوجد الأوزان اللازمة لإنجاز التصانيف التالية: الأشعة (1, 1, 1, 1) و(1-, 1-, 1, 1-) هي أعضاء من الصف (قيمة الخرج تساوي 1)، والأشعة (1-, 1, 1, 1) و(1, 1-, 1-, 1) ليست أعضاء من الصف (الخرج يساوي 1-). استعمل معدل تعليم يساوي 0.5 والقيم الأولية للأوزان تساوي الصفر. استعمل كل أشعة تدريب x كدخل، اختبر استجابة الشبكة.

20.4 اكتب برنامجاً لتصنيف الأحرف من تشكيلات مختلفة باستعمال قاعدة تعليم بيرسبترون. استعمل أكثر ما يمكن من وحدات الخرج حسب ما لديك من أحرف مختلفة في مجموعة التدريب. حوِّل الأحرف إلى الشكل ثنائي القطبية. (قد ترغب بإدخال الأحرف كـ "2" إذا كانت النقطة (pixel) on و"0" إذا كانت off لتسهيل الاختبار مع نماذج ضجيجية بعد التدريب. برنامجك سيطرح 1 من كل مركبة لنماذج الدخل للحصول على شعاع ثنائي القطبية).

1. كرِّر المثال 17 لقيم متعددة للعتبة θ. بعد التدريب –، اختبر لكل قيمة – إمكانية الشبكة على تصنيف النسخ الضجيجية لنماذج التدريب. حاول أن يكون 5، 10، 15، 20 نقطة (pixel) خطأً وبنفس مستويات المعطيات الضائعة. هل للقيم العالية للعتبة أي تأثير بخصوص كيفية اضطراب الشبكة؟ هل توصلت لقيمة العتبة التـــي لا تستطيع الشبكة عندها تعلم كل نماذج التدريب؟

2. جرب بأحرف مختلفة. هل بعض التراكيب أصعب للتعلم من غيرها؟ لماذا؟

21.4 اكتب برنامجاً لتصنيف الأحرف المختلفة باستعمال تعليم قاعدة دلتا. اتبع التوجهات في

التمرين السابق باستثناء القيم المختلفة للعتبة. قارن مقدرة مادالين المدربة على تصنيف دخل ضجيجي مع النتائج في حالة بيرسبترون.

22.4 اكتب برنامجاً لتدريب مادالين لتنفيذ التابع XOR، باستعمال خوارزمية MRI. ما هو تأثير معدلات التعليم المختلفة في الأوزان؟.

شبكات الذاكرة المترافقة
Associative Memory Networks

شبكات الذاكرة المترافقة هي شبكات بسيطة بطبقة أو طبقتين تُخزِّن نماذج يجري استردادها فيما بعد. تشمل هذه الشبكات صنفاً من الشبكات يعرف باسم الذواكر القابلة للعنونة بالمحتوى (Content-addressable memories)، أو أدوات تخزين تسمح باسترداد المعطيات من مفاتيح النموذج المبنية على صفة مميزة للمعطيات المخزنة.

كانت الذواكر المترافقة موضوعاً للبحث الجاد خلال السبعينيات والثمانينيات إلا أن تطبيقاتها ظلّتْ محدودة. رغم ذلك، تشكل هذه الشبكات جزءاً هاماً من البنـــى الأساسية للشبكات العصبونية الصنعية. في هذا الفصل سنشرح الخواص الأساسية لشبكات الذاكرة المترافقة وأنواعها المختلفة والمتنوعة.

1.5 تمهيد

يعتبر التعليم الآلي، أو تخزين الحقائق أو الأشكال والنماذج مع استدلال بسيط أو عدمه، واحداً من أبسط أشكال التعليم، ولكنه الأهم عند الإنسان. ننجز هذا النوع من التعليم عند تخزين جدول الضرب مثلاً. نتعلم بالترافق (الربط) والتخزين في الذاكرة لزوج من الأعداد مع قيمة جدائهما. تُنجز بعدئذٍ عملية استدعاء قيمة الجداء المرافقة باستعمال زوج من الأعداد (الضارب والمضروب) التـــي تعمل كمفاتيح استرداد.

هناك نوع آخر من التعليم النافع هو التخزين والاستدعاء للمعطيات بواسطة محتوى أو درجة التشابه بين نموذج الدخل والنماذج المخزنة، يشار إلى هذا أحياناً بالذاكرة القابلة للعنونة بالمحتوى؛ حيث يكون النموذج المسترد هو النموذج الأكثر شبهاً لتوزيع ما لنموذج الدخل.

يُعرَّف التشابه هنا بقياس ما للمرافقة، مثل التماثل أو المسافة بين نموذجين أو توزيعهما. مثلاً، نستدعي لحناً عندما نسمع فقط بعض النغمات أو نستدعي مكاناً مألوفاً عندما نصادف مكاناً له بعض الملامح المشابهة.

يمكن أن تعمل الشبكات العصبونية الصنعية كذواكر مترافقة عندما تخزن P نموذجاً مختلفاً لتستدعى فيما بعد. عندما يُقدَّم إلى الشبكة المخزنة من قبل نموذج مفتاح دخلها، فإن هذه الشبكة ستعطي على الخرج النموذج المرافق لذاك المفتاح.

الشبكات المترافقة عبارة عن شبكات بسيطة لها بوجه أساسي طبقة وظيفية واحدة كما هو موضح في الشكل (1.5). للشبكة n مدخلاً موزعاً على m وحدة خرج من خلال وصلات مثقّلة بأوزان مختلفة. عموماً، تخزن مصفوفة الوزن n × m أزواجاً مترافقةً P لنماذج الدخل والخرج في شكل تمثيلي موزع $\{(\mathbf{x}^p, \mathbf{y}^p) | p = 1,2,\ldots,P\}$.

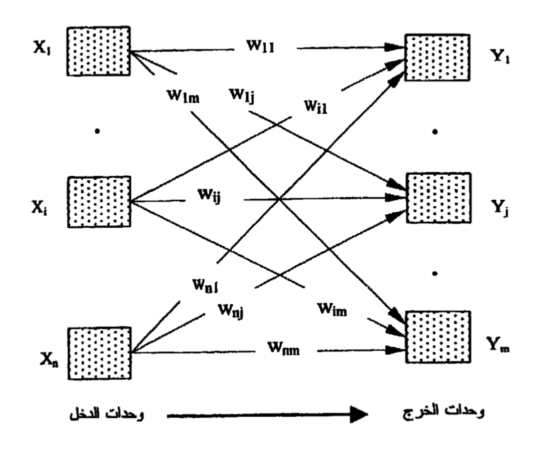
ليكن لدينا P زوجاً لنماذج مترافقة دُرِّبت وخُزِّنت بواسطة الشبكة، إن ما نريده ببساطة من الشبكة هو استرداد النموذج $\mathbf{y}^p$ عندما يكون شعاع دخل الشبكة $\mathbf{x}'$ قريباً من $\mathbf{x}^p$، أي عندما يتحقق:

$$d(\mathbf{x}', \mathbf{x}^p) = \min_q \{d(x', \mathbf{x}^q)\} \qquad (1.5)$$

حيث d (x,y) هي مسافة هامنغ (Hamming) (سنناقش ذلك لاحقاً) أو أية مسافة أخرى قيست بين الشعاعين x وy. كقياس لأداء الشبكة المترافقة، من المألوف استعمال السعة C؛ وهي العدد الأعظمي للنماذج المختلفة بعضها عن بعض التي يمكن تخزينها في الشبكة واستردادها فيما بعد بمعدل خطأ مقبول أصغر من ε. وإذا لم تتحقق المعادلة (1.5) نعلم بأن خطأً ما قد وقع.

تعطى السعة كتابع لحجم الشبكة m عدد الوحدات الوظيفية، كما سنرى فيما بعد. إن جميع شبكات الذاكرة المترافقة لها سعة تخزين محدودة، ومن ثم فلها مقدرة محدودة إلى حد ما. ومع ذلك فإن هذه الشبكات هامة من وجهة النظر العلمية البحتة، ويمكن تنفيذها دون صعوبات كبيرة أو متطلبات مادية معقدة.

يمكننا تنفيذ صنفين من هذه الذواكر المترافقة: يدعى الصنف الأول ذواكر الترافق الذاتي (Autoassociative)، ويدعى الصنف الثاني ذواكر الترافق المغاير (Heteroassociative).



الشكل 1.5: شبكة مترافقة بتغذية أمامية

تستدعـــي ذواكر الترافق الذاتـــي النموذج **y** وهـــو نفس نموذج مفتاح الدخل **x**، أي **x = y**. فعلياً يخزن هذا النوع من الذواكر p نموذجاً وليس نماذج أزواج مترافقة. أما في ذواكر الترافق المغاير، يكون، بوجه عام، النموذج المستدعى **y** شكلاً مختلفاً عن نموذج مفتاح الدخل **x**، أي يكون لدينا $\mathbf{x} \neq \mathbf{y}$. من الواضح في هذه الحالة أنه يجري تخزين أزواج النماذج المترافقة.

يمكن للمرء أن يتساءل عن الفائدة المرجوة من شبكات الترافق الذاتي. بالفعل، لماذا نستدعي نموذج دخل معروف سلفاً؟ ومع أن هذه المهمة تبدو غير نافعة ومضيعة للوقت، فهناك تطبيقات لهذه الذواكر، وخاصة عندما يكون نموذج الدخل ضجيجياً أو غير كامل (لدينا جزء منه فقط) وهو المفتاح الوحيد المتيسر فقط، والنموذج، المخزن من قبل، المستدعى هو الأصل الكامل الخالي من الضجيج.

مثلاً في تطبيقات الرؤية، تحتوي صورة الدخل على أشياء غريبة جزئياً أو قد تكون

مشوهة بسبب الضجيج أو تشويه النظام، والقضية هي استدعاء الأصل أو الأشياء الكاملة بدون ضياع أو تشويه.

وبالمثل يحدث هذا في تطبيقات قنوات الاتصالات، حيث يمكن أن تكون النماذج المبثوثة مشوبة بضجيج القناة، والموضوع الأساسي الأول سيكون تصحيح الخطأ لإعادة الإشارة الأصلية التـــي أرسلت من المنبع (المرسل). في كلتا الحالتين السابقتين، تسترد النماذج الأصلية من نماذج مفتاح غير كاملة من خلال كبت الضجيج، أو بواسطة النموذج الكامل الأصلي وذلك عن طريق إكماله للنموذج الناقص بواسطة الشبكة.

إن أبسط أنواع الشبكات المترافقة هو المرافق الخطي (**Linear associator**). تكون هذه الشبكات أمامية التغذية، حيث يحسب الشعاع المستدعى **y** من شعاع الدخل **x** بواسطة عملية تغذية أمامية وحيدة. النوع الثانـــي للشبكة المترافقة هو نوع التغذية العكسية التكراري، حيث يحسب خرج الشبكة من وحدة أو وحدات متصلة عكسياً مع وحدات الدخل. تستدعي هذه الشبكات نموذج الخرج بعد حساب تكراري حلقي، حيث يمزج الخرج مع الدخل حتـــى تصبح الشبكة مستقرة. ستدرس هذه الشبكات فيما بعد.

ستتصدى الآن فيما تبقى من هذا الفصل لشبكات الطبقة الواحدة، باستثناء الشبكة العصبونية كذاكرة ترافق ثنائية الاتجاه BAM، التـــي لها **n** دخلاً و**m** خرجاً. فيما يخص المترافقات الخطية، يعطى دخل الوحدة رقم j بما يلي:

$$net_j = \sum_{i=1}^{n} w_{ij} x_i = \mathbf{w}_j \cdot \mathbf{x} \quad , j = 1,2,\dots,m \tag{2.5}$$

حيث **x** شعاع الدخل و$\mathbf{w}_j$ هو السطر رقم j من مصفوفة الوزن **W**.

يمكن للمرء أن يتساءل عن الفوائد المكتسبة من استعمال شبكات بطبقات متعددة. تَذْكر من الفصول السابقة أن شبكات الطبقة الواحدة كالبيرسبترون كانت محدودة القدرات الحسابية في مسائل التصنيف الخطي لأشكال ونماذج الدخل.

من ناحية أخرى، كانت الشبكات متعددة الطبقات بتوابع تفعيل غير خطية كمادلين (البيرسبترون متعدد الطبقات) قادرة على تعلم مناطق فصل أكثر تعقيداً. يمكن الإجابة عن هذا التساؤل بأنه لن يكون هناك أية زيادة مكتسبة في سعة التخزين أو القدرة الحسابية

باستعمال مرافقات خطية متعددة الطبقات.

وهذا يتضح من حقيقة أنه في بعض الشبكات، تكون التحويلات الخطية المتعددة مكافئة لتحويل خطي واحد. مثلاً، في الشبكة ثنائية الطبقة ذات شعاع الدخل **x** ومصفوفة الوزن **U**، يعطى خرج الطبقة الأولى بـــ $\mathbf{y} = \mathbf{Ux}$. إذا استعمل خرج الطبقة الأولى **y** كشعاع دخـــل للطبقة الثانية ذات مصفوفة الوزن**V** فسيكـــون خرج الشبكة كـــكل **z** مساوياً لـــ $\mathbf{z} = \mathbf{Vy} = \mathbf{VUx} = \mathbf{Wx}$، حيث $\mathbf{W} = \mathbf{VU}$ مصفوفة وزن الشبكة ككل.

يمكن تعميم هذه النتيجة إلى أي عدد من الطبقات، والقول: **تكافئ التحويلات الخطية المتعددة تحويلاً خطياً واحداً. ومن ثم، لا يجنـــى أي شيء من خلال استعمال طبقات عديدة ذات توابع تفعيل خطية.**

2.5 خوارزميات تعليم الذواكر المترافقة

Learning algorithms for associative memories

لتكن لدينا شبكة بطبقة وحيدة ذات مصفوفة وزن **W**. والسؤال المطروح للبحث هو: كيف تحدد قيم الأوزان لإنجاز تخزين فعال (السعة C كبيرة) واسترداد دقيق؟ سنعتمد ثلاث طرق مستعملة على نطاق واسع في تدريب الشبكات المترافقة، وسنبدأ بقاعدة Hebb.

1.2.5 قاعدة تعليم Hebb

استخدمت قاعدة Hebb البسيطة كنموذج لعدد من الشبكات العصبونية المترافقة. نُذكِّر بأن تعليم Hebb مبنـــي على مبدأ أن التغير في أوزان العصبون متناسب مع إثارة ليفية سابقة ولاحقة، (أي دخل سابق وخرج لاحق).

بكلمات أخرى، إذا كان دخـــل مجموعة مـــن العصبونات هو **x** وتهييـــج الخرج هـــو $\mathbf{y} = \mathbf{y}(\mathbf{x})$ فإن قيم الأوزان الليفية w_{ij} تعطى كما يلي:

$$w_{ij} = \alpha \; x_i \; y_j \qquad (3.5)$$

حيث α ثابت التناسب أو المعايرة، ويمكن كتابة العلاقة السابقة بالشكل المصفوفي التالي:

$$\mathbf{W} = \alpha \, \mathbf{y}(\mathbf{x})^T \qquad (4.5)$$

تخزن مصفوفة الوزن **W** في العلاقة (4.5) فقط زوج نموذج واحد $(\mathbf{x}^p, \mathbf{y}^p)$، هو النموذج

رقم p، ويمكن كتابة:

$$\mathbf{W}^p = \alpha \mathbf{y}^p (\mathbf{x}^p)^T \tag{5.5}$$

ولتخزين نماذج متعددة نركب النماذج الفردية بعضها إلى بعض كما يلي:

$$\mathbf{W} = \sum_{p=1}^{P} \mathbf{W}^p \tag{6.5}$$

في البداية تفرض قيمة الأوزان الأولية مساوية للصفر، وتُحدّث مركبات المصفوفة **W** مع تقدم الزمن باستعمال العلاقة التالية:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha x_i^p y_j^p \quad , \quad p = 1,2,3,\cdots,P \tag{7.5}$$

نجد من العلاقتين (3.5) و(4.5) أن قاعدة Hebb هي بوجه أساسي علاقة تباين مشترك أو ارتباط بين نماذج الدخل والخرج، لذا فإن هذه القاعدة تتطلب معرفة كلا نموذجي الدخل والخرج.

تعطى مصفوفة الأوزان، لتخزين مجموعة من أشعة الدخل s(p) والخرج المنشود t(p) لــ P نموذج دخل معطاة بالشكل التالي:

$$\mathbf{s}(p) = [s_1(p), s_2(p), \ldots, s_n(P)]$$

و

$$\mathbf{t}(p) = [t_1(p), t_2(p), \ldots, t_n(P)]$$

(بالجداءات الخارجية) لأشعة دخل ثنائية بالعلاقة التالية:

$$w_{ij} = \sum_p \big(2s_i(p) - 1\big)\big(2t_j(p) - 1\big) \tag{8.5}$$

وفي حالة أشعة دخل ثنائية القطبية:

$$w_{ij} = \sum_p s_i(p) t_j(p) \tag{9.5}$$

لاحظ أنه عندما نستخدم هذه القاعدة في تعليم الشبكة علينا استخدام الخرج المنشود أو المرغوب به $\mathbf{t}^p$ ليرافق (أو ليرتبط بــ) نماذج الدخل$\mathbf{x}^p$ (أحياناً ندعوها $\mathbf{s}^p$)، بدلاً من الخرج المحسوب $\mathbf{y}^p$ الذي سيختلف عن $\mathbf{t}^p$ لذا سنستمر فيما يلي في استخدام $\mathbf{x}^p$ (أو$\mathbf{s}^p$) لشعاع الدخل في نموذج التدريب رقم p، و$\mathbf{y}^p$ لشعاع الخرج المحسوب بواسطة الشبكة لــ $\mathbf{x}^p$

معطى، وt^p هو الخرج المنشود أو المرغوب به لـــ x^p معطى.

ستطبق هذه القاعدة على شبكات الترافق المتنوعة لاحقاً، لأنها تمتاز باسترداد صحيح تماماً عندما تكون أشعة نماذج الدخل متعامدة مثنى. ولكن أداء الشبكة سيكون أقل عندما تكون أشعة الدخل غير متعامدة. في هذه القاعدة أيضاً تعتمد سعة تخزين الشبكة على درجة الارتباط بين أشعة الدخل.

تعتبر قاعدة Hebb سهلة الحسابات نسبياً لأن كل وزن فيها يعتمد فقط على حدود منفردة (محلية) وليس على أوزان مختلفة. على أية حال فإن الفائدة من هذه القاعدة تعتمد على درجة الارتباط بين نماذج الدخل كما سنرى فيما بعد.

هناك أيضاً مشكلة صعبة في الشكل البسيط لقاعدة تعليم Hebb من العلاقة (3.5)، حيث يبدو جلياً أنه إذا كان معدل التعليم α ثابتاً، فإنه لا يوجد شيء يحد من تزايد قيمة الأوزان ما لم تفرض بعض الشروط المقيدة على قيم نماذج الدخل أو قيم الأوزان. قاد هذا الواقع إلى تعديلات عديدة على قاعدة تعليم Hebb البسيطة.

تشمل هذه التعديلات حد النسيان، أو تشكيل جداءات من شدة الإشارات x وy حول قيمهما المتوسطة، أو تركيب شدة الإشارة مع تغيرات هذه الشدة، وهكذا. اقترحت جميع هذه التغيرات من قبل Tesauro عام 1986[36].

2.2.5 قاعدة دلتا الموسعة

رأينا من قبل أن قاعدة دلتا الأصلية تستعمل توابع تفعيل خطية متماثلة لوحدات الخرج، $\{(\mathbf{x}^p, \mathbf{y}^p) | p = 1,2,..., p\}$ ؛ أو على نحو مكافئ تعدل الأوزان لتقليل الأخطاء المربعة عبر كل نماذج التدريب. باستعمال العلاقة التالية للدخل التركيبـــي المثقل:

$$net_j = \sum_{i=1}^{n} w_{ij} x_i = \mathbf{W}_j \cdot \mathbf{x}$$

فإن خرج الوحدة رقم j يعطى كما يلي:

$$y_j^p = \sum_{i=1}^{n} w_{ij} x_i^p = \mathbf{w}_j . \mathbf{x} \qquad (10.5)$$

يعطـــى في قاعـــدة دلتا الأصليـــة تحديث الأوزان (سنذكر ذلك للمقارنة مع قاعدة دلتا

الموسعة) بالعلاقة التالية:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha(t_j - y_j)x_i \quad i = 1,...,n \;\; ; \;\; j = 1,...,m \qquad (11.5)$$

غالباً ما يعبَّر عن تغير الوزن كما يلي:

$$\Delta w_{ij} = \alpha(t_j - y_j)x_i \qquad (12.5)$$

وهذا يفسر لنا سبب تسمية هذه القاعدة بدلتا (Delta).

إن التعديل الطفيف الذي سنجريه على هذه القاعدة هو استخدام توابع تفعيل تفاضلية كيفية لوحدات الخرج، أي سيكون تحديث الأوزان من وحدة الدخل رقم i إلى وحدة الخرج رقم j كما يلي:

$$\Delta w_{ij} = \alpha(t_j - y_j)\, x_i\, f'(net_j) \qquad (13.5)$$

ما نرغب فيه الآن هو تغيير الأوزان للحصول على أصغر خطأ بين الخرج المحسوب للشبكة والخرج المنشود، عوضاً عن الخطأ بين دخل الشبكة net لوحدات الخرج (ذات التفعيل الخطي) والخرج المنشود في القاعدة الأصلية.

ومن ثم يعطى خطأ الشبكة عبر كل وحدات الخرج m وكل نماذج دخل التدريب P بالعلاقة:

$$E_{tot} = \sum_{p=1}^{P} E^p = \sum_{p=1}^{P}\sum_{j=1}^{m}(t_j^p - \sum_{i=1}^{n} w_{ij}x_i^p)^2 \qquad (14.5)$$

إن تدرج E هو شعاع يتألف من المشتقات الجزئية لـــ E لكل وزن. يعطي هذا الشعاع اتجاه الزيادة السريعة في E، والاتجاه المعاكس سيكون اتجاه النقصان السريع في E، وسيعدّل w_{ij} لتقليل الخطأ الكلي سريعاً باتجاه التدرج السالب؛ $-\frac{\partial E}{w_{ij}}$. يجري ذلك بحل مجموعة المعادلات التالية:

$$\Delta w_{ij} = -\alpha \frac{\partial E_{tot}}{\partial w_{ij}} \qquad (15.5)$$

والحل هو قاعدة دلتا التـــي لها الشكل التالي:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha(t_j^p - y_i^p)x_i^p \qquad (16.5)$$

ويمكن كتابته بالشكل المصفوفي كما يلي:

$$\Delta \mathbf{W} = \alpha \sum_{p=1}^{P} (\mathbf{t}^{p} - \mathbf{W}\mathbf{x}^{p})(\mathbf{x}^{p})^{T} \qquad (17.5)$$

3.2.5 المعكوس الوهمي كتعليم

Pseudo- inverse as learning

هناك طريقة أخرى يمكن تطبيقها لتقليل الخطأ الكلي في نماذج الاسترداد. وفق هذه الطريقة نحصل على مصفوفة الوزن بالحساب المباشر وذلك باستعمال المقلوب الوهمي لمصفوفة الدخل.

تتألف مصفوفة الدخل $\mathbf{X}$ ذات البعد $n \times p$ من P عموداً لأشعة النموذج $\mathbf{x}^{p}$ ذات البعد n. وهكذا سنحسب أولاً مصفوفة المقلوب الوهمي Ж (أثبتنا في الفصل الثالث أن هذه المصفوفة موجودة دائماً حتـــى وإن كانت المصفوفة $\mathbf{X}$ ليست مربعة) لحل مجموعة المعادلات الخطية التالية:

$$\mathbf{W}\,\mathbf{X} = \mathbf{T} \qquad (18.5)$$

حيث $\mathbf{T}$ مصفوفة أشعة الهدف المترافقة مع أشعة مصفوفة الدخل $\mathbf{X}$.

يعطى الحل المصفوفي (راجع الفقرة 3.2.3 من الفصل الثالث)كما يلي:

$$\mathbf{W} = \mathbf{T}\,\text{Ж}$$

$$\mathrm{W} = \lim_{\lambda \to 0} \left[\mathbf{T}\mathbf{X}^{\mathrm{T}} (\mathbf{X}\mathbf{X}^{\mathrm{T}} + \lambda \mathbf{I})^{-1} \right] = \mathbf{T}\,\text{Ж} \qquad (19.5)$$

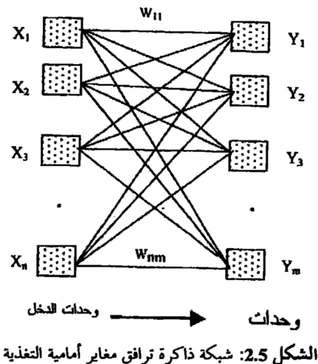
تعطي المعادلة (19.5) حلّ مربعات أصغرياً بمفهوم أخطاء الاسترداد، وهي نفسها قاعدة دلتا عندما تكون أشعة الدخل عبارة عن مجموعة مستقلة خطياً.

عندما يكون هناك أكثر من حل، نختار الحل الذي يكون فيه مجموع المربعات أصغرياً. بالطبع عندما يكون للمصفوفة $\mathbf{X}$ مقلوب عادي $\mathbf{X}^{-1}$ نكتب $\text{Ж} = \mathbf{X}^{-1}$.

3.5 شبكات ذاكرة الترافق المغاير أمامية التغذية

Feedforeward heteroassociative memory networks

تخـــزن شبكاتُ الترافـــق المغاير الأمامية التغذية أزواج النماذج **(x, t)**، حيث **x ≠ t** في



الشكل 2.5: شبكة ذاكرة ترافق مغاير أمامية التغذية

في الحالة العامة. إذا كان لدينا P نموذجاً لزوج مترافق $(\mathbf{x}^p,\mathbf{t}^p)$، حيث $p = 1, 2, .., P$، و$\mathbf{x}^p$ هو شعاع ببعد n و$\mathbf{t}^p$ شعاع ببعد m، فالمسألة تؤول إلى تصميم شبكة عصبونية كذاكرة ترافق مغاير موضحة في الشكل (2.5)، تعطي نموذج الاسترداد $\mathbf{y}^p$، حيث $\mathbf{y}^p = \mathbf{t}^p$، عندما يطبق على دخل الشبكة النموذج $\mathbf{x}^p$.

بتطبيق قاعدة Hebb نستطيع مباشرة حساب مصفوفة الارتباط $\mathbf{W}^p$ لكـــي يكون الزوج $(\mathbf{x}^p,\mathbf{t}^p)$ مترافقاً:

$$\mathbf{W}^p = \mathbf{t}^p(\mathbf{x}^p)^T \qquad (20.5)$$

حيث العنصر رقم ij من هذه المصفوفة يساوي: $w_{ij} = t_j^p x_i^p$.

لاسترداد نموذج مفرد $\mathbf{t}^p$ من $\mathbf{W}^p$ لنموذج الدخل $\mathbf{x}^p$، نستعمل مداخل خطية لوحدات الشبكة (أي ذات ترافق خطي المعادلة (2.5) أو بالشكل المصفوفي $\mathbf{W}^p\mathbf{x}^p$ وتوابع تفعيل التماثل $f(net) = net$ للمخارج. في هذه الحالة، يمكن استرداد $\mathbf{t}^p$ مباشرة إذا كان نموذج

الدخل $\mathbf{x}^p$ معيارياً، أي إذا كان $\|\mathbf{x}^p\|^2$ فإن:

$$\mathbf{W}^p\,\mathbf{x}^p = \mathbf{t}^p(\mathbf{x}^p)^T.\mathbf{x}^p = \mathbf{t}^p \qquad (21.5)$$

بالطبع نريد تخزين أكثر من نموذج واحد، ومن ثم:

$$\mathbf{W}\sum_{p=1}^{P}\mathbf{W}^p, P = 1,2,..,\mathbf{P} \qquad (22.5)$$

لاسترداد النموذج $\mathbf{t}^k$ عندما يكون دخل الشبكة $\mathbf{x}^k$ المرافق سيكون لدينا:

$$\mathbf{W}\,\mathbf{x}^k = (\sum_{p=1}^{P}\mathbf{t}^p(\mathbf{x}^p)^T).\mathbf{x}^k$$

$$= \mathbf{t}^k(\mathbf{x}^k)^T.\mathbf{x}^k + \left(\sum_{p\neq k}^{P}\mathbf{t}^p(\mathbf{x}^p)^T\right)\mathbf{x}^k \qquad (23.5)$$

لاحظ أن الحد الأول في طرف المعادلة (23.5) الأيمن يساوي $\mathbf{t}^k$ إذا كان الجداء السلمي $(\mathbf{x}^k)^T.\mathbf{x}^k = 1$، وهذا سيكون إذا كان الشعاع $\mathbf{x}$ معيارياً. $\|\mathbf{x}^p\| = 1$ وكذلك فإن الحد الثانـي من الطرف الأيمن للمعادلة (23.5) يساوي الصفر إذا كانت الأشعة $\mathbf{X}$ متعامدة مثنـى بعضها مع بعض؛ وهذا يعنـي أنه في حالة $p \neq k$:

$$(\mathbf{x}^p)^T.\mathbf{x}^k = 0$$

بكلمات أخرى، ستعطي الشبكة استدعاء تاماً إذا كانت نماذج الدخل متعامدة مثنـى.

بالطبع العدد الأعظمي للنماذج التـي يمكن تخزينها في الشبكة سيكون m، حيث $m \leq n$، لأن العدد الأعظمي للأشعة المتعامدة مثنـى في فراغ ببعد n يساوي n. حتـى عندما لا تكون أشعة الدخل متعامدة، فإن الاستدعاء الدقيق يبقى ممكناً إذا كانت أشعة الدخل معيارية (نظيمها يساوي الواحد) والحد الثانـي في الطرف الأيمن "حد التداخل" للمعادلة (23.5) كان بقيمة صغيرة مقارنةً مع الحد $\mathbf{t}^k$، وهذا سيحصل عندما تكون أشعة الدخل تقريباً متعامدة أو مرتبطة ارتباطاً ضعيفاً بعضها مع بعض.

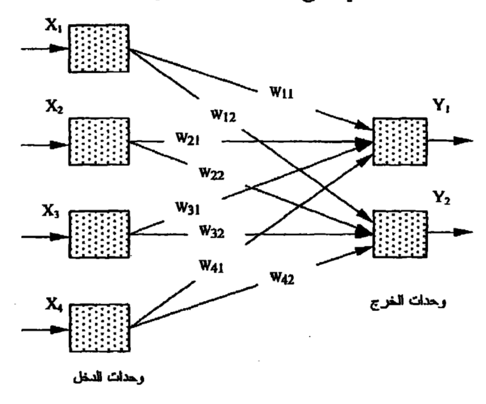
عندما تكون أشعة الدخل ارتباطية، فإن حد التداخل سيضع قيوداً على سعة التخزين ويمكن أن تحدث استردادات كبيرة يكون فيها $\mathbf{y}^k \neq \mathbf{t}^k$.

بالطبع يمكن استخدام طرق التعليم المختلفة المشروحة في الفقرة 2.5 السابقة لتعيين شعاع الوزن بنتائج محسنة في الدقة و/أو السعة عندما لا تكون أشعة الدخل متعامدة. وسنعطي

بعض الأمثلة لتوضيح المفاهيم المذكورة آنفاً.

1.3.5 تطبيقات شبكات ذواكر الترافق المغاير

سنعتمد في جميع هذه الأمثلة شبكة ذاكرة الترافق المغاير الأمامية التغذية الوحيدة الطبقة بأربع وحدات دخل ووحدتـــي خرج، الموضحة في الشكل (3.5).



الشكل 3.5: شبكة ذاكرة الترافق المغاير

مثال 1:

شبكة ترافق مغاير باستعمال قاعدة Hebb، ومعطيات ثنائية، المطلوب تدريب الشبكة الموضحة في الشكل (5.3) بقاعدة Hebb لتخزين ترافقات أشعة (سطر الدخل $s = (s_1, s_2, s_3, s_4)$ (غير المتعامدة مثنـــى) إلى أشعـــة (سطر) الخرج $t = (t_1, t_2)$ التالية:

	مركبات نماذج الدخل					مركبات نماذج الخرج	
	s_1	s_2	s_3	s_4		t_1	t_2
s^1	1	0	0	0	t^1	1	0
s^2	1	1	0	0	t^2	1	0
s^3	0	0	0	1	t^3	0	1
s^4	0	0	1	1	t^4	0	1

على أي حال، فإن قيم الخرج المختارة مرتبطة بأشعة الدخل بأسلوب خاص بسيط، والتداخل بين شعاع الدخل الأول وشعاع الدخل الثانـــي لا يسبب أي صعوبات (لأن قيم أشعة الخرج لها ستكون متساوية).

يعرّف التدريب المنجز بواسطة قاعدة Hebb كما يلي:

$$w_{ij}^{new} = w_{ij}^{old} + s_i t_j$$

$$\Delta w_{ij} = s_i t_j$$

وتابع التفعيل الثنائي لوحدة الخرج رقم j يعطى بواسطة:

$$y_j = f(net_j) = \begin{cases} 1 & net_j > 0 \\ 0 & net_j \leq 0 \end{cases}$$

لتعيين قيمة مصفوفة الوزن سنستخدم الجداء الخارجي، وهو ببساطة جداء شعاع دخل التدريب المكتوب كشعاع عمود (ببعد n × 1) وشعاع الخرج المكتوب كشعاع سطر (ببعد 1 × m).

وستكون مصفوفة الوزن لتخزين كل من أزواج التدريب كما يلي:

$$\mathbf{W}^1 = (\mathbf{s}^1)^T \mathbf{t}^1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{W}^2 = (\mathbf{s}^2)^T \mathbf{t}^2 = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{W}^3 = (\mathbf{s}^3)^T \mathbf{t}^3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{W}^4 = (\mathbf{s}^4)^T \mathbf{t}^4 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} [0 \quad 1] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

وسنحصل على مصفوفة الوزن المستخدمة لتخزين أزواج النماذج المترافقة الأربعة من خلال جمع مصفوفات الأشعة منفردة لتخزين كل زوج نموذج على حدة:

$$\mathbf{W} = \mathbf{W}^1 + \mathbf{W}^2 + \mathbf{W}^3 + \mathbf{W}^4$$

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\mathbf{W} = \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}$$

والآن سنقوم باختبار مقدرة الشبكة على استرداد نموذج الخرج لكل مداخل التدريب.

$\mathbf{y}^1 = f\,(\mathbf{s}^1\mathbf{W})$

$$\mathbf{y}^1 = f([1 \quad 0 \quad 0 \quad 0] \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}) = f([2 \quad 0]) = [1 \quad 0]$$

$\mathbf{y}^2 = f\,(\mathbf{s}^2\mathbf{W})$

$$\mathbf{y}^2 = f([1 \quad 1 \quad 0 \quad 0] \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix}) = f([3 \quad 0]) = [1 \quad 0]$$

$\mathbf{y}^3 = f(\mathbf{s}^3\mathbf{W})$

$$\mathbf{y}^3 = f([0 \quad 0 \quad 0 \quad 1]\begin{bmatrix}2 & 0\\1 & 0\\0 & 1\\0 & 2\end{bmatrix}) = f([0 \quad 2]) = [0 \quad 1]$$

$$\mathbf{y}^4 = f(\mathbf{s}^4\mathbf{W})$$

$$\mathbf{y}^4 = f([0 \quad 0 \quad 1 \quad 1]\begin{bmatrix}2 & 0\\1 & 0\\0 & 1\\0 & 2\end{bmatrix}) = f([0 \quad 3]) = [0 \quad 1]$$

نلاحظ أن الشبكة استردت النماذج الأربعة استرداداً صحيحاً، وكان هذا متوقعاً لأن أشعة دخل التدريب كانت مستقلة خطياً وإن لم تكن هذه الأشعة متعامدة. نهيب بالقارئ إعادة هذا المثال على أشعة دخل غير مستقلة خطياً ومقارنة إنجاز الشبكة مع النتائج السابقة.

والآن سنقوم باختبار هذه الشبكة في استرداد نموذج للخرج عند تطبيق أي نموذج مشابه لنماذج دخل التدريب. ليكن شعاع الاختبار $s` = (0, 1, 0, 0)$ الذي يختلف عن شعاع التدريب $s2 = (1, 1, 0, 0)$ فقط في أول مركبة. سيكون لدينا:

$$\mathbf{s}'.\mathbf{W} = [0,1,0,0].\mathbf{W} = [1 \quad 0] \xrightarrow{f(net)} \mathbf{y}' = [1 \quad 0]$$

وهكذا نجد أن الشبكة أرفقت أحد نماذج الخرج المعروفة بنموذج الاختبار وهذا ما نرغب به.

وسنحاول الآن اختبار الشبكة بالنموذج $s` = (0, 1, 1, 0)$ الذي يختلف عن كل نماذج دخل التدريب على الأقل في مركبتين. سيكون لدينا:

$$\mathbf{s}'.\mathbf{W} = [0,1,1,0].\mathbf{W} = [1 \quad 1] \xrightarrow{f(net)} \mathbf{y}' = [1 \quad 1]$$

إن خرج الشبكة ليس واحداً من نماذج الخرج المترافقة التـــي دربت عليها، وبكلمات أخرى: لم تستطع الشبكة تمييز هذا النموذج. نلاحظ أن هذا النموذج يختلف عن نموذج دخل التدريب $s^2 = (1, 1, 0, 0)$ في المركبة الأولى والثالثة، ولذلك يوجد خطآن في نموذج الدخل، وهذا ما سيجعل الشبكة عاجزة عن تمييز النموذج. وهذا ليس غريباً لأنه يمكن أن ينظر إلى نموذج الاختبار على أنه مكوّن أيضاً من نموذج التدريب $s^4 = (0, 0, 1, 1)$ مع أخطاء في المركبة الثانية والرابعة.

مثال 2:

شبكة ترافق مغاير باستعمال قاعدة Hebb، ومعطيات ثنائية القطبية، إن التمثيل الثنائي القطبية لنماذجنا مفضل حسابياً بوجه عام عن التمثيل الثنائي، وسيعطي أداءً أفضل للشبكة. لتوضيح ذلك دعنا ندرس أشعة التدريب المترافقة للشبكة في المثال السابق ممثلة بالشكل الثنائي القطبية كما يلي:

	مركبات نماذج الدخل					مركبات نماذج الخرج	
	s_1	s_2	s_3	s_4		t_1	t_2
$\mathbf{s}^1$	1	−1	−1	−1	$\mathbf{t}^1$	1	−1
$\mathbf{s}^2$	1	1	−1	−1	$\mathbf{t}^2$	1	−1
$\mathbf{s}^3$	−1	−1	−1	1	$\mathbf{t}^3$	−1	1
$\mathbf{s}^4$	−1	−1	1	1	$\mathbf{t}^4$	−1	1

لتخزين مجموعة أزواج الأشعة الثنائية القطبية $(\mathbf{s}^p, \mathbf{t}^p)$ لكل نموذج تدريب p، حيث

$$\mathbf{t}^p = (t_1^p, t_2^p, \ldots, t_m^p) \quad \text{و} \quad \mathbf{s}^p = (s_1^p, s_2^p, \ldots, s_n^p)$$

يمكن حساب قيم عناصر مصفوفة الوزن $\{w_{ij}$ من العلاقة التالية:

$$w_{ij} = \sum_{p=1}^{P} s_i^p \, t_j^p$$

وسنوضح عملية إيجاد المصفوفات باستعمال الجداء الخارجي. ففي حالة زوج النموذج الأول تعطى مصفوفة الوزن بما يلي:

$$\mathbf{W}^1 = (\mathbf{s}^1)^T \mathbf{t}^1 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

وفي حالة الزوج الثانـــي تعطى بـــ:

$$\mathbf{W}^2 = (\mathbf{s}^2)^T \mathbf{t}^2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

وفي حالة الزوج الثالث تعطى بـــ:

$$\mathbf{W}^3 = (\mathbf{s}^3)^T \mathbf{t}^3 = \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ 1 & -1 \\ -1 & 1 \end{bmatrix}$$

وفي حالة الزوج الرابع تعطى بـــ:

$$\mathbf{W}^4 = (\mathbf{s}^4)^T \mathbf{t}^4 = \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & -1 \\ -1 & 1 \\ -1 & 1 \end{bmatrix}$$

وستكون مصفوفة الوزن لتخزين الأزواج الأربعة هي مجموع مصفوفات الأزواج اللازمة لتخزين كل زوج نموذج على حدة:

$$\mathbf{W} = \sum_{p=1}^{P} \mathbf{W}^P \text{ ،} \qquad \mathbf{W} = \begin{bmatrix} 4 & -4 \\ 2 & -2 \\ -2 & 2 \\ -4 & 4 \end{bmatrix}$$

إن إحدى الفوائد الحسابية للتمثيل الثنائي القطبية هي قدرته على إعطاء طريقة بسيطة للتعبير عن مستويين مختلفين للضجيج الذي يمكن أن يطبق على مداخل التدريب للحصول على مداخل اختبار للشبكة. سنشير إلى هذين النوعين من الضجيج "بالمعطيات الضائعة" و"بالأخطاء".

مثلاً، إذا كان كل نموذج من نماذج التدريب الأصلية متتالية من الاستجابات نعم أو لا، فإن "المعطيات الضائعة" ستوافق استجابة **لا أعرف**، على حين أن "الخطأ" سيمثل الاستجابة **نعم** عندما تكون الاستجابة الصحيحة **لا** والعكس بالعكس. بالتمثيل الثنائي القطبية، **نعم** تمثل بـــ +1 و**لا** تمثل بـــ -1 و**لا أعرف** تمثل بـــالصفر (وهذا يبدو جلياً عندما نمثل مركبات الضجيج في تطبيقات تعرّف الأشكال أو الأحرف).

رأينا في التمثيل الثنائي أن الشبكة كانت عاجزة على تعرّف نماذج الاختبار التـــي اعتراهـــا التشويه أو الضجيج في مركبتيـــن، وكذلك الأمر في التمثيل الثنائـــي القطبية، حيث لا تستطيع الشبكة إعطاء الجواب الصحيح في حالة شعاع دخل اختبار مثل $s`=(-1, 1, 1, -1)$ مكوّن من شعاع مخزن بمركبتين خطأ كما يلي:

$$[-1 \quad 1 \quad 1 \quad -1].\mathbf{W} = [0 \quad 0]$$

$$f([0 \quad 0]) = [0 \quad 0]$$

وهذه الاستجابة خاطئة. على أية حال تستطيع الشبكة أن تعطي استجابة صحيحة عندما تعطى شعاع دخل مكوّن من شعاع مخزن بمركبتين ضائعتين، فمثلاً في حالة شعاع الاختبار $s`= (0, 1, 0, -1)$ الذي يعتبر شكلاً من شعاع التدريب $s = (1, 1, -1, -1)$ حيث المركبتان الأولى والثالثة ضائعتان (عوضاً عن الخطأ). سيكون لدينا:

$$[0 \quad 1 \quad 0 \quad -1].\mathbf{W} = [6 \quad -6]$$

$$f([6 \quad -6]) = [1 \quad -1]$$

هذه الاستجابة صحيحة في حالة الشعاع المخزن $s = (1, 1, -1, -1)$.

2.3.5 تطبيقات شبكة الترافق المغاير في تعرف الأشكال

مثال 3:

تطبيق آخر لشبكة الترافق لأحرف مترافقة من بين تشكيلات مختلفة. دربت الشبكة في هذه الحالة باستعمال قاعدة Hebb على ثلاثة أزواج لأشعة مترافقة مبينة في الشكل (4.5). الأشعة x لها $7\times9= 63$ مركبة، والأشعة y لها $5\times3=15$ مركبة. كل مركبة مثلناها بالرمز "#" إذا كانت من الشكل وبالرمز "0" إذا لم تكن من الشكل، وجميع هذه الأشعة تمثل نماذج ثنائية البعد.

```
000#000           #####00          00###00
000#000           #0000#0          0#000#0
00#0#00   0#0     #00000#   ##0    #00000#   ###
00#0#00   #0#     #0000#0   #0#    #000000   #00
00###00   ###     #####00   ##0    #000000   #00
0#000#0   #0#     #0000#0   #0#    #000000   #00
0#000#0   #0#     #00000#   ##0    #00000#   ###
#00000#           #0000#0          0#000#0
#00000#           #####00          00###00
دخل       خرج     دخل       خرج    دخل       خرج
```

الشكل 4.5: نماذج تدريب لتعرّف الأحرف على استعمال شبكة ترافق مغاير

سنقوم الآن بتحويل كل نموذج ثنائي إلى شعاع رقمي بقيم ثنائية القطبية، وذلك بتمثيل الرمز "#" الذي يمثل نقطة مضاءة في الحرف بالرقم 1 والرمز "0" الذي يمثل نقطة غير مضاءة في الحرف بالرقم 1–، وتجري القراءة سطراً سطراً من أعلى سطر ومن أقصى اليسار إلى اليمين، مثلاً النموذج التالي:

```
0#0
#0#
###
#0#
#0#
```

الشكل 5.5: نموذج ثنائي البعد للحرف A

يصبح شعاعاً بخمس عشرة مركبة كما يلي:

$$[-1\ \ 1\ \ -1,\ 1\ \ -1\ \ 1,\ 1\ \ 1\ \ 1,\ 1\ \ -1\ \ 1,\ 1\ \ -1\ \ 1]$$

استعملت الشبكة بعد تدريبها مع نماذج دخل ضجيجية حيث كانت نسخة مشوهة عن نماذج دخل التدريب. والنتائج معروضة في الشكلين (6.5) و(7.5). حيث يأخذ الضجيج على الشكل التالي:

الرمز "X" هو الآن 1 ويمثل خطأ ضجيج

الرمز "*" هو الآن 0 ويمثل خطأ ضجيج.

دخل	خرج	دخل	خرج	دخل	خرج
x00#000 000#000 00#0#00 00#0#00 00###0x 0#000#0 0#0x0#0 #00x00# #00x00#	0#0 #0# ### #0# #0#	x00#00x 000#000 x0#0#00 00#0#00 x0###0x 0#000#0 0#0x0#x #00x00# #00x00#	0#0 #0# ### #0# #0#	x00#00x 000#0x0 x0#0#00 00#0#00 x0###0x 0xx00#0 0#0x0#x #00x00# #00x00#	0#0 #0# ### #0# #0#

دخل	خرج	دخل	خرج	دخل	خرج
x00#00x 000#0x0 x0#0#0x 00#0#0x x0###0x 0#x00#0 0#0x0#x #0xx00# #00x0x#	0#0 #0# ### #0# #0#	000#000 000*000 00#0#00 00*0#00 00#*#00 0#000*0 0#000*0 #00000# #00000#	0#0 #0# ### #0# #0#	000#000 000*000 00*0#00 00*0#00 00#**00 0#000*0 0#000*0 *00000# *00000#	0#0 #0# ### #0# #0#

دخل	خرج	دخل	خرج	دخل	خرج
00x#0x0 0x0#x00 x0#x#0x 00*0#x0 0x##*00 x#0x0#x 0#x00*0 x00x0# *00x00*	0#0 #0# [illegible] #0# #0#	*#*##00 #x000*0 #x00x0# *00x0*0 ##*##x0 #000x*0 *00x00* #0x00*0 #*##*00	##0 #0# ##0 #0# ##0	x0#*#0x 0#x00*0 #x00x0# *00x00x #0x00x0 #0x0x00 *00x00* 0#x00*0 0x##*00	### #00 #00 #00 [illegible]

الشكل 6.5: استجابة شبكة الترافق المغاير لنسخ مختلفة من الضجيج للنموذج A

```
00x#0x0              *#*##00              x0#*#0x
0x0#x00              #x000*0              0#x00*0
x0#x#0x     0#0      #x00x0#     ##0      #x00x0#     ###
00*0#x0     #0#      *00x0*0     #0#      *00x00x     #00
0x##*00     ###      ##*##x0     ##0      #0x00x0     #00
x#0x0#x     #0#      #000x*0     #0#      #0x0x00     #00
0#x00*0     #0#      *00x00*     ##0      *00x00*     ###
#x00x0#              #0x00*0              0#x00*0
*00x00*              #*##*00              0x##*00
```

دخل خرج دخل خرج دخل خرج

الشكل 7.5: استجابة شبكة ترافق مغاير للنماذج A, B, C بأخذ حوالي 1/3 المركبات كضجيج

يظهر الشكل (7.5) أن الشبكة العصبونية يمكن أن تميز الأحرف الصغيرة المخزنة فيها حتـــى عندما تعطى نماذج دخل تمثل أحرف تدريب كبيرة تبلغ نسبة الضجيج فيها 33 % تقريباً.

4.5 شبكات ذاكرة الترافق الذاتـــي الأمامية التغذية

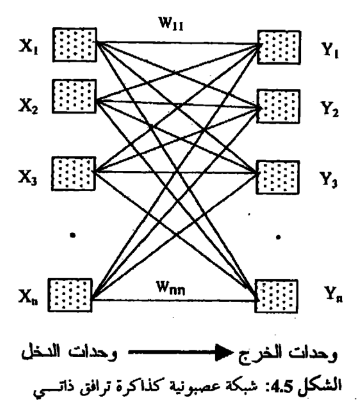
Feedforward autoassociative memory networks

تعتبر شبكات الترافق الذاتـــي حالة خاصة من شبكات الترافق المغاير التـــي جرى عرضها في المقطع السابق. تستخدم هذه الشبكات لتخزين أزواج أشعة دخل التدريب وأشعة الخرج المنشود (x, t) المتماثلة؛ أي x = t ومن ثم عدد وحدات الدخل n يساوي عدد وحدات الخرج m، وبنية الشبكة تتلخص بطبقة واحدة أمامية التغذية كما هو موضح في الشكل (5.4). تسمى عملية تدريب الشبكة بتخزين الأشعة، ويمكن أن تكون الأشعة ثنائية أو ثنائية القطبية.

كما في حالة شبكات الترافق المغاير، يمكن استعمال أية طريقة من طرق التعليم الثلاثة المشروحة سابقاً لتدريب شبكات الترافق الذاتي. يُستردّ الشعاع المخزن من دخل ضجيجي جزئي (غير كامل)، أو مشوه، إذا كان هذا الدخل مشابهاً للشعاع المخزن بقدر كاف. ويحكم على إنجاز الشبكة من خلال مقدرتها على استرداد النموذج الكامل المخزن عندما يكون الدخل ضجيجياً. وسيكون إنجاز الشبكة أفضل في التمثيل الثنائي القطبية منه في التمثيل الثنائي.

غالباً ما تأخذ قيم الأوزان القطرية (الأوزان على الوصلات بين مركبة نموذج الدخل والمركبة الموافقة لها في نموذج الخرج) قيماً افتراضية مساوية للصفر في مصفوفة الأوزان لهذه الشبكات.

إن وضع قيم هذه الأوزان مساوياً للصفر سيحسن مقدرة الشبكة على التعميم، وخاصة إذا كان هناك أكثر من شعاع واحد مخزناً فيها (Szu عام 1989[37])، كما يمكن أن نزيد معقولية الشبكة بيولوجياً (Anderson عام 1972[38]). إن إعطاء قيمة الصفر لهذه الأوزان ضروري لتوسيـــع الحالة التكرارية (Hopfield عام 1982[10])، أو عند استخدام قاعدة دلتا لمنع التدريب من إعطاء مصفوفة واحدية للأوزان (McClelland & Rumelhart عام 1986[20]).



الشكل 4.5: شبكة عصبونية كذاكرة ترافق ذاتـــي

1.4.5 تطبيقات شبكة الترافق الذاتـــي في التخزين والاسترداد

مثال 4:

سنوضح عملية التخزين في شبكة الترافق الذاتـــي ومن ثم الاسترداد للنماذج المخزنة. سنعتمد قاعدة Hebb لتدريب الشبكة والتمثيل الثنائي القطبية لأزواج دخل التدريب والخرج المنشود، فيكون تابع تفعيل وحدة الخرج رقم j كما يلي:

$$y_j = f(net_j) = \begin{cases} 1 & if\ net_j > 0 \\ -1 & if\ net_j \le 0 \end{cases} \qquad (22.5)$$
$$j = 1,2,\ldots,n$$

ستكون أشعة تدريب الدخل $\mathbf{s} = (s_1,s_2,s_3,s_4,s_5)$، حيث $\mathbf{s} = \mathbf{t}$، كما يلي:

$$\mathbf{s}^1 = [+1\ +1\ +1\ +1\ -1]\ ,\ \mathbf{s}^2 = [-1\ +1\ -1\ +1\ -1]$$
$$\mathbf{s}^3 = [+1\ +1\ -1\ -1\ +1]\ ,\ \mathbf{s}^4 = [+1\ +1\ +1\ -1\ -1]$$

يمكن حساب مصفوفة الوزن لتخزين كل أزواج النماذج المترافقة باستعمال الجداء الخارجي (كما فعلنا تماماً في شبكات الترافق المغاير ولكن لدينا الآن $\mathbf{s}^p = \mathbf{t}^p$ لكل نموذج تدريب) كما يلي:

من أجل زوج النموذج الأول:

$$\mathbf{W}^1 = (\mathbf{s}^1)^T \mathbf{t}^1 = \begin{bmatrix} +1 \\ +1 \\ +1 \\ +1 \\ -1 \end{bmatrix} \begin{bmatrix} +1 & +1 & +1 & +1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} +1 & +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 & -1 \\ +1 & +1 & +1 & +1 & -1 \\ -1 & -1 & -1 & -1 & +1 \end{bmatrix}$$

$$\mathbf{W}^2 = (\mathbf{s}^2)^T \mathbf{t}^2 = \begin{bmatrix} -1 \\ +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} \begin{bmatrix} -1 & +1 & -1 & +1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} +1 & -1 & +1 & -1 & +1 \\ -1 & +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 \\ -1 & +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 & +1 \end{bmatrix}$$

$$\mathbf{W}^3 = (\mathbf{s}^3)^T \mathbf{t}^3 = \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \\ +1 \end{bmatrix} \begin{bmatrix} +1 & +1 & -1 & -1 & +1 \end{bmatrix}$$

$$= \begin{bmatrix} +1 & +1 & -1 & -1 & +1 \\ +1 & +1 & -1 & -1 & +1 \\ -1 & -1 & +1 & +1 & -1 \\ -1 & -1 & +1 & +1 & -1 \\ +1 & +1 & -1 & -1 & +1 \end{bmatrix}$$

$$\mathbf{W}^4 = (\mathbf{s}^4)^T \mathbf{t}^4 = \begin{bmatrix} +1 \\ +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} +1 & +1 & +1 & -1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} +1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & -1 & -1 \\ +1 & +1 & +1 & -1 & -1 \\ -1 & -1 & -1 & +1 & +1 \\ -1 & -1 & -1 & +1 & +1 \end{bmatrix}$$

ومصفوفة الوزن **W** التـــي ستخزن جميع أشعة التدريب الأربعة ستكون:

$$\mathbf{W} = \sum_{P=1}^{4} = \mathbf{W}^P \begin{bmatrix} 4 & 2 & 2 & -2 & 0 \\ 2 & 4 & 0 & 0 & -2 \\ 2 & 0 & 4 & 0 & -2 \\ -2 & 0 & 0 & 4 & -2 \\ 0 & -2 & -2 & -2 & 4 \end{bmatrix}$$

يُستردّ كل نموذج مخزن على النحو التالي:

$$\mathbf{s}^1\mathbf{W} = [+6+8+8+4-10] \Rightarrow$$

$$\mathbf{y}^1 = f([+6 \quad +8 \quad +8 \quad +4 \quad -10]) = [+1 \quad +1 \quad +1 \quad +1 \quad -1]$$

وبنفس الطريقة تُسترد جميع الأزواج المخزنة في المصفوفة والنتائج ستكون كما يلي:

$$s^2W=[-6 \quad +4 \quad -4 \quad +8 \quad -6] \Rightarrow$$
$$y^2=f([-6 \quad +4 \quad -4 \quad +8 \quad -6])=[-1 \quad +1 \quad -1 \quad +1 \quad -1]$$
$$s^3W=[6 \quad +4 \quad -4 \quad -8 \quad +6] \Rightarrow$$
$$y^3=f([+6 \quad +4 \quad -4 \quad -8 \quad +6])=[+1 \quad +1 \quad -1 \quad -1 \quad +1]$$
$$s^4W=[+10 \quad +8 \quad +8 \quad -4 \quad -6] \Rightarrow$$
$$y^4=f([+10 \quad +8 \quad +8 \quad -4 \quad -6])=[+1 \quad +1 \quad +1 \quad -1 \quad -1]$$

وهكذا نجد أن الشبكة أعطت استجابة صحيحة لكل نماذج التدريب.

إن الشبكة قادرة أيضاً على إعطاء الاستجابة الصحيحة في حالة نماذج دخل ضجيجية مختلفة قليلاً عن النماذج المخزنة. وكما ذكرنا من قبل يُشوّه النموذج للحصول على نموذج اختبار بطريقتين: "الأخطاء" في المعطيات و"فقدان" المعطيات. سنعتبر من أجل "الأخطاء" أن التغير في المركبات سيكون من +1 إلى −1 أو بالعكس، أما تمثيل المعطيات الضائعة، فسنعبر عنها بإعطاء قيمة الصفر للمركبة بدلاً من +1 أو −1.

سنقوم الآن باختبار الشبكة بمجموعة من الأشعة الضجيجية s' المكوّنة من الشعاع الأصلي $s=(1, 1, 1\ -1)$، المخزن في مصفوفة الوزن **W**، مع خطأ في مركبة واحدة فقط.

$$W=\begin{bmatrix} 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & 1 & 1 & -1 \\ -1 & -1 & -1 & 1 \end{bmatrix}$$

$$s' \cdot W=[-1 \quad 1 \quad 1 \quad -1] \cdot W=[2 \quad 2 \quad 2 \quad -2] \xrightarrow{f(net)} [1 \quad 1 \quad 1 \quad -1]$$
$$s' \cdot W=[1 \quad -1 \quad 1 \quad -1] \cdot W=[2 \quad 2 \quad 2 \quad -2] \xrightarrow{f(net)} [1 \quad 1 \quad 1 \quad -1]$$
$$s' \cdot W=[1 \quad 1 \quad -1 \quad -1] \cdot W=[2 \quad 2 \quad 2 \quad -2] \xrightarrow{f(net)} [1 \quad 1 \quad 1 \quad -1]$$
$$s' \cdot W=[1 \quad 1 \quad 1 \quad 1] \cdot W=[2 \quad 2 \quad 2 \quad -2] \xrightarrow{f(net)} [1 \quad 1 \quad 1 \quad -1]$$

لاحظ أن جميع استجابات الشبكة كانت صحيحة، ويستطيع القارئ التأكد من أن الشبكة ستستجيب استجابة صحيحة لأشعة الاختبار s' المكوّنة من الشعاع المخزن مع مركبة واحدة فقط ضائعة (قيمتها تساوي الصفر) هي: $(1, 1, 1, 0)$ $(1,1,0,-1)$، $(1,0,1,-1)$، $(0,1,1,-1)$.

ستكون استجابة الشبكة عموماً أفضل في حالة المعطيات الضائعة منها في حالة الخطأ، وهذا

ليس غريباً لأن الأشعة مع المعطيات الضائعة تكون أقرب(رياضياً وبالحدس) إلى نماذج التدريب من الأشعة مع المعطيات الخاطئة. مثلاً، ستعطي الشبكة استجابة صحيحة لأشعة الاختبار `s المكوّنة من الشعاع الأصلي المخزن (1, 1, 1, 1) = s بضياع مركبتين كما يلي:

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[0\quad 0\quad 1\quad -1]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[0\quad 1\quad 0\quad -1]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[0\quad 1\quad 1\quad 0]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[1\quad 0\quad 0\quad -1]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[1\quad 0\quad 1\quad 0]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[1\quad 1\quad 0\quad 0]\cdot\mathbf{W}=[2\quad 2\quad 2\quad -2]\xrightarrow{f(net)}[1\quad 1\quad 1\quad -1]$$

يدل هذا على أن الشبكة تعرّفت جميع نماذج الاختبار، وأدركت أنها نسخة مشوهة عن النموذج الأصلي s المعروف لديها. لنر الآن كيف ستكون استجابة الشبكة لنماذج الاختبار المكونة من أخطاء في النموذج الأصلي. ليكن شعاع الاختبار التالي (1–, 1, 1–, 1–)=`s من الشعاع الأصلي المخزن (1–, 1, 1, 1) = s بمركبتين خطأ (المركبة الأولى والثانية). ستكون استجابة الشبكة:

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}=[-1\quad -1\quad 1\quad -1]\cdot\mathbf{W}=[0\quad 0\quad 0\quad 0]\xrightarrow{f(net)}[0\quad 0\quad 0\quad 0]$$

هذا يعني أن الشبكة لم تميز شعاع الدخل هذا. غالباً ما تؤخذ مصفوفة الأوزا ن بعناصر صفرية على قطرها الرئيسي؛ وهذا يعني عدم وجود وصلات بين وحدات الدخل ومثيلاتها المقابلة لها من وحدات الخرج. وستصبح مصفوفة الوزن السابقة كما يلي:

$$\mathbf{W}_0=\begin{bmatrix}0 & 1 & 1 & -1\\ 1 & 0 & 1 & -1\\ 1 & 1 & 0 & -1\\ -1 & -1 & -1 & 0\end{bmatrix}$$

وسنرى ما هي استجابة الشبكة في حالة شعاع الاختبار السابق (1–, 1, 1–, 1–)=`s.

$$\mathbf{s}^{\backprime}\cdot\mathbf{W}_0=[-1\quad -1\quad 1\quad -1]\cdot\mathbf{W}_0=[1\quad 1\quad -1\quad 1]\xrightarrow{f(net)}[1\quad 1\quad -1\quad 1]$$

لاتزال استجابة الشبكة خاطئة. الآن سنقوم باختبار الشبكة بالأشعة `S المكوّنة من الشعاع الأصلي المخزن (1–, 1, 1, 1) = s بضياع مركبتين فقط.

$$\mathbf{s}'\cdot\mathbf{W_0}=[0\ \ 0\ \ 1\ \ -1]\cdot\mathbf{W_0}=[2\ \ 2\ \ 1\ \ -1]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$
$$\mathbf{s}'\cdot\mathbf{W_0}=[0\ \ 1\ \ 0\ \ -1]\cdot\mathbf{W_0}=[2\ \ 1\ \ 2\ \ -1]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$
$$\mathbf{s}'\cdot\mathbf{W_0}=[0\ \ 1\ \ 1\ \ 0]\cdot\mathbf{W_0}=[2\ \ 1\ \ 1\ \ -2]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$
$$\mathbf{s}'\cdot\mathbf{W_0}=[1\ \ 0\ \ 0\ \ -1]\cdot\mathbf{W_0}=[1\ \ 2\ \ 2\ \ -1]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$
$$\mathbf{s}'\cdot\mathbf{W_0}=[1\ \ 0\ \ 1\ \ 0]\cdot\mathbf{W_0}=[1\ \ 2\ \ 1\ \ -2]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$
$$\mathbf{s}'\cdot\mathbf{W_0}=[1\ \ 1\ \ 0\ \ 0]\cdot\mathbf{W_0}=[1\ \ 1\ \ 2\ \ -2]\xrightarrow{f(net)}[1\ \ 1\ \ 1\ \ -1]$$

وهكذا نجد أن الشبكة قد تعرّفت جميعَ أشعة الدخل وميزتها جيداً.

مثال 5:

سعة تخزين شبكة الترافق الذاتـــي لأشعة متعامدة، الاعتبار الهام للشبكات العصبونية كذاكرة مترافقة هو عدد النماذج أو أزواج النماذج التـــي تستطيع تخزينها قبل أن تبدأ الشبكة بالنسيان. تستطيع شبكة الترافق الذاتـــي تخزين أكثر من شعاع واحد، وذلك بجمع مصفوفات الأوزان لكل شعاع. مثلاً، إذا استعملت مصفوفة الوزن $\mathbf{W}_1$ لتخزين الشعاع (1, 1– 1, –1) والمصفوفة $\mathbf{W}_2$ لتخزين الشعاع (1– ,1 ,1 ,1–)، فإن مصفوفة الوزن $\mathbf{W}=\mathbf{W}_1+\mathbf{W}_2$ ستستعمل لتخزين كلا الشعاعين السابقين (المتعامدين).

من المألوف إعطاء أوزان القطر الرئيسي قيمة الصفر، لأننا إذا لم نفعل ذلك فإن أوزان القطر الرئيسي ستهيمن ـــ ومن ثم ـــ ستميل الشبكة إلى إعطاء شعاع الدخل عوضاً عن الشعاع المخزن.

$$\overset{\mathbf{W}_1}{\begin{bmatrix} 0 & 1 & -1 & -1 \\ 1 & 0 & -1 & -1 \\ -1 & -1 & 0 & 1 \\ -1 & -1 & 1 & 0 \end{bmatrix}} + \overset{\mathbf{W}_2}{\begin{bmatrix} 0 & -1 & -1 & 1 \\ -1 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 \\ 0 & -1 & -1 & 0 \end{bmatrix}} = \overset{\mathbf{W}}{\begin{bmatrix} 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \\ -2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \end{bmatrix}}$$

ويمكن للقارئ أن يتأكد بسهولة أن الشبكة تستطيع تعرّف كلا الشعاعين.

وهكذا فإن سعة تخزين الشبكة هي عدد الأشعة التـــي تستطيع تخزينها. تستطيع شبكة الترافق الذاتـــي بأربعة عقد تخزين ثلاثة أشعة متعامدة فقط، ولا تستطيع الشبكة بأربع عقد تخزين الشعاع الرابع حتـــى وإن كانت جميع الأشعة متعامدة مثنـــى.

لتكن W_3 مصفوفة الوزن لتخزين الشعاع (1 ,1-, 1 ,1-). عندئذ ستكون مصفوفة الوزن الكلية القادرة على تخزين كل الأشعة الثلاثة هي مجموع المصفوفات W_1 وW_2 وW_3:

$$\underset{W_1 + W_2}{\begin{bmatrix} 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \\ -2 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \end{bmatrix}} + \underset{W_3}{\begin{bmatrix} 0 & -1 & 1 & -1 \\ -1 & 0 & -1 & 1 \\ 1 & -1 & 0 & -1 \\ -1 & 1 & -1 & 0 \end{bmatrix}} = \underset{W_1 + W_2 + W_3}{\begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}}$$

والتـــي ستُصنِّف تصنيفاً صحيحاً كل واحد من الأشعة الثلاثة التـــي دربت عليها. نحاول تخزين الشعاع الرابع (1 ,1 ,1 ,1) المتعامد مع بقية الأشعة مثنـــى وبمصفوفة الوزن W_4.

$$\underset{W_1 + W_2 + W_3}{\begin{bmatrix} 0 & -1 & -1 & -1 \\ -1 & 0 & -1 & -1 \\ -1 & -1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}} + \underset{W_4}{\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}} = \underset{W}{\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}$$

التـــي لا تتعرّف أيَّ شعاع.

لذا يمكن تلخيص النتائج التالية:

تعتمد سعة شبكة الترافق الذاتـــي على عدد مركبات الأشعة المخزنة وعلى العلاقة بين هذه الأشعة، وستزداد سعة التخزين كلما ازداد عدد الأشعة المخزنة المتعامدة مثنـــى. ويمكننا إثبات أنه يمكن تخزين n −1 شعاعاً ثنائي القطبية متعامدة مثنـــى باستعمال مجموع مصفوفات أوزان الجداءات الخارجية (بأوزان صفرية على القطر الرئيسي)، حيث n عدد مركبات كل شعاع، لشبكة ترافق ذاتـــي لها n وحدة دخل وخرج، **ولكن أية محاولة لتخزين n شعاعاً متعامداً مثنـــى سيؤدي إلى عدم استجابة الشبكة لأي شعاع مخزن.** ما سبق يكمل دراستنا عن الشبكات المترافقة الأمامية التغذية.

5.5 شبكات Hopfield

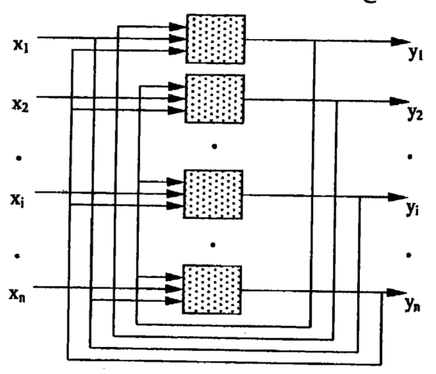
هي شبكة ترافق ذاتـــي تكرارية مشابهة للشبكات الموصوفة سابقاً، اقترحت هذه الشبكة

من قبل Hopfield عام 1982[10]، وعام 1984[39]. إن شبكة هوبفيلد (Hopfield) هي شبكة تكرارية بطبقة واحدة ذات مصفوفة أوزان متناظرة عناصر القطر الرئيسي فيها كلها أصفار.

ففي شبكة هوبفيلد الموضحة في الشكل (5.5) بـ n وحدة، ستكون مصفوفة الأوزان **W** فيها $w_{ij} = w_{ji}$ و $w_{ii} = 0$ في حالة $i, j = 1,2,...,n$.

بسبب بساطة بنية هذه الشبكة، يمكن أن يفاجأ المرء بكيفية إعطاء خرج مختلف عن الدخل. يجري ذلك من خلال العملية التكرارية للشبكة كما سنشرحه لاحقاً.

تخزِّن شبكات هوبفيلد عدداً ما P من النماذج الأولية التي تسمى بجواذب النقطة الثابتة (fixed-point attractor). ويحدَّد المحل الهندسي للجواذب بواسطة الأوزان **W**. ويمكن أن تخصَّص النماذج المخزنة بواسطة حساب مباشر، مثلاً بتعليم Hebb، أو يمكن أن تعلم من خلال تدرج الهبوط كقاعدة دلتا.



الشكل 5.5: شبكة Hopfield متقطعة

عندما تصبح الشبكة مدربة على P نموذجاً أولياً مثل $x^1, x^2, ..., x^P$، تكون قادرة

على الاسترداد المترافق. مثلاً، لاستدعاء النموذج $\mathbf{x}^k$ تعمل الشبكة تكرارياً على تغذية إشارات خرجها عكسياً إلى المداخل (الخرج من الوحدة i يتصل مع جميع مداخل الوحدات ماعدا مدخل الوحدة i) مكررة ذلك عند كل لحظة تحديث زمنـــي t حتـــى تصبح الشبكة مستقرة تماماً.

في البداية تغذى إشارة الدخل $\mathbf{x}(0)$ إلى مداخل الوحدات في اللحظة الزمنية $t = 0$. ومن ثم تحسب المخارج بواسطة الوحدات. في الأنظمة الرقمية يعين خرج الشبكة بواسطة معادلات الفروق التالية:

$$x_i(t+1) = \text{sgn}(\sum_{j=1}^{n} w_{ij} x_j(t) - \theta) \quad ,i = 1,2,...,n \qquad (24.5)$$

حيث العتبة $\theta \geq 0$. وسنستعمل تابع تفعيل ثنائي القطبية $\{1-,+1\}$ مع $x_i \in \{1-,+1\}$

$$\text{sgn}(x) = \begin{cases} +1 & if\ x > 0 \\ -1 & if\ x < 0 \end{cases} \qquad (25.5)$$

و بالطبع في حالة $x = 0$ يكون لدينا $x_i(t+1) = x_i(t)$.

ستكون البداية مع الدخل $\mathbf{x}$، ومن ثم ستحسب المخارج من التابع $\text{sgn}(x)$ (25.5) وتغذى عكسياً من الدخل وفق خوارزمية تحديث ما. تحسب المخارج الجديدة وتغذى عكسياً ثانية من المداخل في الخطوة الزمنية التالية، وهكذا تتكرر العملية حتـــى تستقر الشبكة على نقطة ثابتة موافقة للنموذج المُعلَّم.

تعرّف مصفوفة الأوزان لتخزين مجموعة من أشعة الدخل $\mathbf{s(p)}$ والخرج المنشود $\mathbf{t(p)}$ لـــ P نموذج دخل، حيث:

$$\mathbf{s(p)} = [\mathbf{s}_1(\mathbf{p}), \mathbf{s}_2(\mathbf{p}), ..., \mathbf{s}_n(\mathbf{P})]$$

و

$$\mathbf{t(p)} = [\mathbf{t}_1(\mathbf{p}), \mathbf{t}_2(\mathbf{p}), ..., \mathbf{t}_n(\mathbf{P})]$$

في حالة أشعة دخل ثنائية كما يلي:

$$\mathbf{w}_{ij} = \sum_{\mathbf{p}} (2\mathbf{s}_i(\mathbf{p}) - 1)(2\mathbf{t}_j(\mathbf{p}) - 1) \qquad (26.5)$$

وفي حالة أشعة دخل ثنائية القطبية:

$$w_{ij} = \sum_{p} s_i(p) t_j(p) \qquad (27.5)$$

يمكن أن تكون خوارزمية التحديث متزامنة، أو غير متزامنة أو مزيجاً من الحالتين معاً. ففي حالة التحديث المتزامن، تحسب المخارج $x_i(t+1), i = 1,2,...,n$ (المعادلة (25.5)) كمجموعة في لحظة زمنية واحدة قبل أن تغذى المخارج عكسياً إلى المداخل.

وفي التحديث غير المتزامن، تحسب المخارج $x_i(t+1)$ بالتتالي بترتيب ما أو بالانسجام مع توزيع احتمالي ما، ومن ثم تغذى عكسياً إلى المداخل بعد كل عملية تحديث. مثلاً، قد تكون أوزان الوحدات معدلة تعديلاً متتالياً حسب ترتيب أدلتها.

أما في الطور المتزامن وغير المتزامن، فتحدَّث مجموعات جزئية من الوحدات على نحو متزامن، وبعدئذ تُحدَّث كل مجموعة باستعمال خوارزمية اختيار غير متزامنة. خلال الاستدعاء، يمكن أن تصل شبكة هوبفيلد عامة في النهاية إلى واحدة من الحالتين التاليتين:

1. حلقة، ففي حالة قيمة كبيرة بمقدار كافٍ لـــ t، ودور ثابت $T > 1$ فإن $\mathbf{x}(t) = \mathbf{x}(t+1)$.

2. نقطة ثابتة معرفة بواسطة $\mathbf{x}(t+1) = \mathbf{x}(t)$ في حالة قيمة كبيرة بمقدار كافٍ لـــ t.

للعمل كذاكرة مترافقة، يجب أن تتقارب الشبكة لنقطة ثابتة $\mathbf{x}^j$ بحيث تكون قريبة لشعاع الدخل $\mathbf{x}(0)$ بعد عدد محدود من التكرارات. وسيكون ذلك دائماً عندما تكون مصفوفة الأوزان $\mathbf{W}$ متناظرة.

1.5.5 خصائص تابع الطاقة

Energy function characterization

واحد من الأمور الهامة لشبكات هوبفيلد هو تمييز حالة الشبكة بتابع الطاقة. ولما كانت مصفوفة الأوزان في هذه الشبكات متناظرة فإنه من الممكن تعريف تابع الطاقة كما يلي:

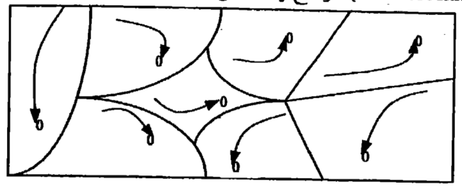
$$E = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} x_i x_j \qquad (28.5)$$

(وهذا التابع مشابه لتابع الطاقة الذي يميز المواد المغناطيسية في الفيزياء).

استطاع هوبفيلد أن يثبت أنه كلما تطور نظام الشبكة العصبونية الصنعية وفقاً لديناميكيته فإن الطاقة يجب أن تصل في النهاية إلى الحالة المستقرة التـــي لا يمكن لتابع الطاقة عندها المعرف E أن يزداد بعد كل تحديث (Hopfield عام 1982[10])، ويجب أن يتناقص

أو يبقى ثابتاً. وبسبب وجود عدد محدد من الحالات، فإن الشبكة يجب أن تتقارب في النهاية إلى الأصغر المحلي. إن الطاقة الصغرى توافق جاذب النقطة الثابتة؛ أي النماذج المخزنة، وتعيِّن حالة النظام عند التقارب نموذج الخرج.

فعلياً، تعتمد حالة النظام على الحالة الأولية للشبكة ومصفوفة الأوزان **W**. فإذا كانت **W** ثابتة، فإن كل الحالات الأولية ضمن مسافة معينة من نقطة الجذب تكوّن ما يسمى بحوض الجذب (Basin of Attraction) الموضح في الشكل (6.5).



الشكل 6.5: جواذب النقطة-الثابتة وحوض الجذب

من كل حالة أولية يعينها نموذج الدخل، سيتطور النظام بالحركة أسفل سطح تابع الطاقة حتـــى يصل إلى الأصغر المحلي. لكي نثبت أن تابع الطاقة (العلاقة (28.5)) لا يزداد أبداً بعد الوصول إلى نقطة الأصغر المحلي، سنرمز لتغير الطاقة من E إلى E' بـ $\Delta E = E' - E$ بعد أن تكون وحدة ما قد تحدثت ولتكن الوحدة رقم k. إن الطاقة تبقى ثابتة أي $\Delta E = 0$ إذا لم تتغير حالة الوحدة. من ناحية أخرى، سنفرض أن الوحدة k قد تغيرت حالتها من x_k إلى x'_k بحيث $x'_k = -x_k$. لذا ستُحذف كل الحدود في ΔE غير المتغيرة فيصبح ΔE كما يلي:

$$\Delta E = -\frac{1}{2}\sum_{j\neq k} w_{kj} x'_k x_j + \frac{1}{2}\sum_{j\neq k} w_{kj} x_k \, x_j$$

$$\Delta E = -\frac{1}{2} x'_k \sum_{j\neq k} w_{kj} x_j + \frac{1}{2} x_k \sum_{j\neq k} w_{kj} x_j \tag{29.5}$$

$$= x_k \sum_{j\neq k} w_{kj} x_j$$

حيث استنتج الحد الأخير من حقيقة أن $x'_k = -x_k$. لاحظ أن الحد الأخير في علاقة ΔE سالب دائماً، لأن حد المجموع و x_k هما بإشارتين متعاكستين بالفرض (تذكر أن تغير

الحالة إلى E' كان مفروضاً، وأن $x'_k = \text{sgn}\left(\sum_j w_{kj}x_j\right) = -x_k$ مع كون $w_{ii} = 0$). وهكذا سيكون دائماً ΔE متناقصاً أو سيبقى ثابتاً عند كل تحديث.

أخيراً يجب أن يصل النظام إلى حالة الطاقة الصغرى المحلية لأن E محددة تحديداً ضئيلاً من الأسفل (لها قيمة عتبة صغرى محددة بالأوزان)؛ لأن لكل **x** لدينا المتراجحة:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_jx_i \geq -\sum_{i=1}^{n}\sum_{j=1}^{n}\left|w_{ij}\right| \tag{30.5}$$

و w_{ij} جميعها قيم محددة. كما ذكرنا من قبل توافق الطاقة الصغرى النموذج المخزن. إن المناقشة السابقة كانت لشبكة هوبفيلد المتقطعة، ولكن هذه الشبكة عُمِّمت لتعمل في الزمن المستمر وتعطي قيم خرج مستمرة.

2.5.5 شبكات هوبفيلد المستمرة

النسخة المستمرة لشبكة هوبفيلد هي تعميم مباشر للشبكة المتقطعة من خلال استعمال توابع تفعيل بقيم مستمرة بدلاً من توابع التفعيل الثنائية القطبية (أو الثنائية). تُستعمل عادة توابع تفعيل الظل القطعي $\tanh(x)$ أو Sigmoid، ويجري تعديل ديناميكية الشبكة خلال الزمن المستمر. لهذا ستصف مجموعة غير خطية من المعادلات التفاضلية سلوك الشبكة. هذه المعادلات أعطيت من قبل GrossbergوCohen عام 1983[40] وHopield عام 1984[39]، وهي:

$$\tau_i \frac{dx_i}{dt} = -x_i + f\left(\sum_{j=1}^{n} w_{ij}x_j\right) \quad ,i = 1,2,\cdots,n \tag{31.5}$$

حيث τ_i هو ثابت زمني، و $f(x)$ هو تابع التفعيل غير الخطي:

$$f(x) = \frac{1}{1-\exp(-x)} \tag{32.5}$$

بوجه عام، تعرض ديناميكية النظام الموصوفة في المعادلات العامة السابقة (27.5) ثلاثة أنواع من السلوك هي:

1. التقارب إلى نقطة ثابتة
2. سلوك اهتزازي

3. سلوك فوضوي(راجع الفقرة 7.3 من الفصل الثالث).

ولما كانت مصفوفة الأوزان متناظرة، فمن المفترض أن تتقارب الشبكة إلى نقطة ثابتة. برهن Cohen-Grossberg هذا التقارب وسنأتـــي على شرحه لاحقاً. يكون النظام في حالة توازن عندما يصل إلى النقطة الثابتة، أي إن التفعيل لا يتغير، ويكون:

$$\frac{dx_i}{dt} = 0 \quad \text{لكل قيم } i \tag{33.5}$$

وستثبت مخارج الوحدات بعدئذ عند:

$$x_i = f(\sum_{i=1}^{n} w_{ij} x_j) \quad \text{لكل قيم } i \tag{34.5}$$

حيث x_i ذات قيم حقيقية محصورة بين [0, 1]. تعطى الصيغ المكافئة لديناميكية النظام بحدود الاشتقاق للمداخل net لكل وحدة بدلاً من المخارج، ويعطى المدخل التركيـــي لكل وحدة بالمعادلات التالية (يجب ألاّ ننسى أن $w_{ii} = 0$ في جميع المعادلات الخاصة بشبكة Hopfield):

$$net_i = \sum_{j=1}^{n} w_{ij} x_j \quad ,i = 1,2,...,n \tag{35.5}$$

يشبه نظام المعادلات التفاضلية في هذه الحالة نظام المعادلات (31.5). بعد وضع net_j محل x_i سنحصل على المعادلة التالية:

$$\tau_i \frac{dnet_i}{dt} = -net_i + \sum_{j=1}^{n} w_{ij} f(net_j) \quad ,i = 1,2,...,n \tag{36.5}$$

عند التقارب، تأخذ هذه المعادلات شكل الحل المبين في المعادلة (34.5) المسماة تابع طاقة Lyapunov:

$$net_i = \sum_{j=1}^{n} w_{ij} f(net_j) \quad ,i = 1,2,...,n \tag{37.5}$$

استطاع هوبفيلد عام 1984[39] أن يعطي تابع الطاقة للحالة المستمرة:

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} x_j x_i + \sum_{i=1}^{n}\int_{0}^{x_i} f^{-1}(x)d(x) \tag{38.5}$$

يمكن أن نرى أن E تتناقص، ومن ثمَ فهي كما في الحالة المتقطعة.

باشتقاق المعادلة (38.5) بالنسبة إلى الزمن والقيام بالتعويضات والاختصارات المناسبة، يمكن أن نثبت أن $dE/dt = 0$. استخدمت شبكات هوبفيلد المستمرة بداية في تطبيقات الاستمثال كمسألة البائع الجوال والجدولة وغيرها، واستخدمت في تطبيقات أخرى مثل معالجة الصور (Bilbro عام 1988[41]) والتحكم (Tsutsumi عام 1988[28]).

عندما نصمِّم شبكة لحل مسألة في الاستمثال، فإن قيم الأوزان والانحياز تستعمل للتعبير عن الشروط المقيدة للمسألة المعالجة، ويكون إيجاد مجموعة المعادلات المقيدة لحل هذه الشبكة هو مفتاح الحل الناجح. سيكون عادة حل الشبكة ليس استمثالياً تماماً ولكن قريب منه، وهذا مفيد في حالات كثيرة.

لقد أُعطِيَ جوابٌ جزئي عن مشكلة تقارب الأنظمة غير الخطية بتغذية عكسية في النظرية العامة التـــي ترتبط مع الاستقرار الكلي لصنف كبير من الأنظمة الديناميكية. وقد وضع هذه النظرية Cohen-Grossberg عام 1983[40]، وطُبِّقت على بنـــى الشبكات العصبونية الصنعية الديناميكية، وعُمِّمتْ حتـــى وسعت المجال التطبيقي من قبل Kosko (a,b) عام 1988[42],[43].

أما فيما يخص سعة تخزين هذه الشبكة، فإن هوبفيلد أثبت تجريبياً أن عدد النماذج الثنائية التـــي تستطيع أن تخزنها وتستدعيها هذه الشبكة بدقة مقبولة يعطى تقريبيًّا بالمقدار:

$$P \approx 0.15n \qquad (39.5)$$

حيث n عدد العصبونات في الشبكة. لقد أنجز Abo-Mostafa و St Jacques عام 1985 [44]تحليلاً نظرياً مفصلاً عن سعة المعلومات لشبكة هوبفيلد. وفي حالة شبكة مشابهة وباستعمال نماذج ثنائية القطبية، أثبت كل من McEliece ,Posner ,Rodemich ,venkatesh عام 1987 [45] أن سعة التخزين تساوي تقريباً:

$$P \approx \frac{n}{2\log_2 n} \qquad (40.5)$$

أخيراً، هناك عدد من التعديلات التـــي أجريت على شبكة هوبفيلد المتقطعة. فقد استخدم هوبفيلد ـــ بوجه أساسي ـــ تفعيلات ثنائية بدون دخل خارجي بعد أول خطوة زمنية (Hopfield عام 1982[10])، ومن ثم أصبح الدخل الخارجي موجوداً باستمرار طوال

تنفيذ الخوارزمية (Hopfieldعام 1984[39])، واستعمل Hecht-Nielsen عام 1990 [46] تفعيلات ثنائية القطبية ولكن بدون دخل خارجي.

3.5.5 نظرية Cohen-Grossberg

تثبت نظرية Cohen-Grossberg تقارب مسارات صنف من الأنظمة الديناميكية غير الخطية؛ فهي تنص على أن أي تابع ديناميكي غير خطي يعطى بالمعادلة:

$$\frac{dx_i}{dt} = f_i(x_i)\left[g_i(x_i) + \sum_{j=1} w_{ij}h_j(x_j)\right] \quad ,i=1,2,\dots,n \qquad (41.5)$$

حيث:

ـــ المصفوفة $\mathbf{W} = (W_{ij})$ متناظرة مع عناصر ثابتة غير سالبة.

ـــ التابع $f_i(x)$ لجميع القيم $x \geq 0$.

ـــ التابع $g_i(x)$ لجميع القيم $x > 0$.

ـــ التابع $f_i(x) > 0$ لجميع القيم $x > 0$ والتابع $h_i(x) \geq 0$ لجميع القيم x.

ـــ التابع $h_i(x)$ مستمر وقابل للاشتقاق وغير متناقص لجميع القيم $x \geq 0$.

ـــ في حالة $i = 1,2,\dots,n$ لدينا:

$$\lim_{x\to\infty} \sup[g_i(x) - w_{ii}h_i(x)] < 0$$

ـــ $\lim_{x\to 0+} g_i(x) = \infty$

و

$$\int_0^{\varepsilon} \frac{dx}{f_i(x)} = \infty \quad ,\varepsilon > 0$$

بعدئذ تقترب كل المسارات المقبولة من مجموعة غير متغيرة ضخمة M محتواة ضمن المجموعة E:

$$E = \{y \in R^n: dv(y)/dt = 0 \; ; y \geq 0\}$$

$$dV/dt = -\sum_{i=1}^{n} f_i \; h_i'\left[g_i - \sum_{k=1}^{n} w_{ik}h_k\right]^2$$

لاحظ أنه إذا ازداد كل تابع h_i ازدياداً حاداً، فإن المجموعة E تتألف من كل نقاط

التوازن لنظام المعادلات (41.5).

يتطلب برهان نظرية Cohen-Grossberg أن نظام المعادلات المعطى بـ (41.5) يُعرِّف تابع Lyapunov ويحقق شروطاً أخرى معينة (مبدأ اللاتغيرية لـ LaSalle وشروط نظرية Sard`s).

ويجب على المرء لاستعمال النظرية أن يكون قادراً على التعبير عن ديناميكية النظام بالصيغة المعطاه بالمعادلات (41.5) وأن يتأكد له تحقّق كل شروط النظرية.

مثال 6:

اختبار شبكة Hopfield المتقطعة باستخدام قاعدة Hebb، ليكن لدينا الشعاع $s = (1, 1, 1, 0)$ المخزن في الشبكة. وسنحاول اختبار الشبكة بالشعاع $s` = (0, 0, 1, 0)$ المكوّن من مركبتين خطأ، هما الأولى والثانية، في الشعاع المخزن **s**.

يجري تحديث تفعيلات الوحدات بترتيب كيفي، ولكن في هذا المثال سنفترض أن ترتيب التحديث كما يلي: Y_2, Y_3, Y_4, Y_1، قيمة العتبة تساوي الصفر وتابع تفعيلات الوحدات هو:

$$y_i = f(net_i) = \begin{cases} 1 & net_i > 0 \\ y_i & net = 0 \\ 0 & net_i < 0 \end{cases}$$

1. سنفترض أن مصفوفة الأوزان لتخزين النماذج هي:

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & 1 & 0 & -1 \\ -1 & -1 & -1 & 0 \end{bmatrix}$$

2. شعاع الدخل الخارجي هو $\mathbf{x} = (0, 0, 1, 0)$ لهذا الشعاع.
3. الخرج في اللحظة $t = 0$، $\mathbf{y} = (0, 0, 1, 0)$.
4. سنختار الوحدة Y_1 لتحديث تفعيلها:

$$net_1 = x_1 + \sum_j y_j w_{j1} = 0 + 1 = 1 \quad —$$

— $net_1 > 0 \rightarrow y_1 = 1$

— الخرج في اللحظة الزمنية $t = 1$، $\mathbf{y} = (1, 0, 1, 0)$

5. اختيار الوحدة Y_4 لتحديث تفعيلها:

— $net_4 = x_4 + \sum_j y_j w_{j4} = 0 - 2 = -2$

— $net_4 < 0 \rightarrow y_4 = 0$

— الخرج في اللحظة الزمنية $t = 2$، $\mathbf{y} = (1, 0, 1, 0)$

6. اختيار الوحدة Y_3 لتحديث تفعيلها:

— $\text{net}_3 = \text{x}_3 + \sum_j \text{y}_j \text{w}_{j3} = 1 + 1 = 2$

— $net_3 > 0 \Rightarrow y_3 = 1$

— الخرج في اللحظة الزمنية $t = 3$، $\mathbf{y} = (1, 0, 1, 0)$

7. اختيار الوحدة Y_2 لتحديث تفعيلها:

— $\text{net}_2 = \text{x}_2 + \sum_j \text{y}_j \text{w}_{j2} = 0 + 2 = 2$

— $net_2 > 0 \rightarrow \text{y}_2 = 1$

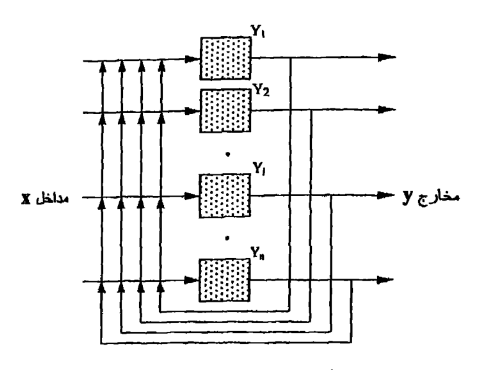
— الخرج في اللحظة الزمنية $t = 4$، $\mathbf{y} = (1, 1, 1, 0)$

8. اختبار شرط التقارب.

ولما كان بعض التفعيلات قد تغير خلال دور التحديث، فلا بد من تكرار دور تحديث آخر على الأقل، ويستطيع القارئ التحقّق أنه لن يحدث أي تغير في التفعيلات لأي وحدة. وهذا يعنـــي أن الشبكة قد تقاربت إلى النموذج المخزن.

6.5 حالة دماغ في صندوق Brain State-in-a-Box (BSB)

قدم Anderson وزملاؤه عام 1977[47] شبكة عصبونية صنعية بصفتها ذاكرة ترافق ذاتـــي تكرارية أسموها حالة دماغ في صندوق. يوضح الشكل (7.5) بنية هذه الشبكة.



الشكل7.5: شبكة حالة دماغ في صندوق

أتت هذه التسمية من كون حالة النظام محصورة ضمن مكعب واحدي (+1,−1). تتألف الشبكة من طبقة واحدة من n وحدة، فهي تشبه كثيراً شبكة هوبفيلد المتقطعة من حيث البنية ومبدأ العمل باستثناء أنه لا توجد قيود مفروضة على مصفوفة الأوزان. وبوجه خاص، يمكن أن يكون للوحدات تغذية عكسية ذاتية، أي $w_{ii} \neq 0$. كما يمكن أن تكون بعض الوصلات ملغاة نهائياً، أي $w_{ij} = 0$ ، لبعض i وj.

يمكن استعمال تعليم Hebb أو تدرج الهبوط (قاعدة دلتا) لتعديل الأوزان في الشبكة BSB. عند استعمال قاعدة تعليم Hebb، تمنع الأوزان من النمو إلى ما بعد حد "الإشباع"، ويعيَّن هذا الحد بوضع تفعيل الوحدات عند القيم +1 أو −1. طبعاً هذا ينجز بواسطة تابع تفعيل الوحدات التالي:

$$x_j(t+1) = \begin{cases} +1 & net_j(t) > +1 \\ x_j & |net_j| \leq 1 \\ -1 & net_j < -1 \end{cases} \qquad (42.5)$$

حيث $x_j(t+1)$ هو خرج الوحدة j في اللحظة t +1 و$net_j(t)$ هو دخل الوحدة j في اللحظة t:

$$net_j(t) = x_j(t) + \alpha \sum_{i=1}^{n} w_{ij} x_i(t) \quad (43.5)$$

حيث α عامل التعليم.

في البداية، تكون الشبكة متوضعة في أي مكان ضمن المكعب الواحدي (ثنائي القطبية). ولما كان النظام يتطور بالانسجام مع ديناميكيته، فإن حالته تتحرك تكرارياً باتجاه أحد الأوجه حيث تصبح مقيدة هناك (ممنوعة من النمو). توافق الأضلاع جواذب النظام الديناميكي غير الخطي، أي النماذج المخزنة.

عندما تُستعمل قاعدة تعليم Hebb، تعطى الأوزانُ قيمةً أولية كيفية صغيرة $w_{ij} << 1$، ويقدم شعاع النموذج $\mathbf{x}$ إلى دخل الشبكة. بعدئذ يحسب تفعيل الوحدات من التابع (42.5) ومن ثم تنتشر المخارج إلى كل الوحدات من خلال وصلات التغذية العكسية، وهكذا تتكرر العملية حتـــى تصبح الشبكة مستقرة. وتُحدَّث الأوزان وفقاً لقاعدة Hebb فور استقرار الشبكة:

$$w_{ij}^{new} = w_{ij}^{old} + \beta x_i x_j \quad (44.5)$$

حيث x_i و x_j هما لتدريب الشبكة فإن الأوزان تعدل وفقاً لـــ:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha(t_i - x_i)x_j \quad (45.5)$$

حيث t_i الخرج المنشود للوحدة i، و x_i الخرج الفعلي المحسوب للوحدة i، و x_j خرج الوحدة j المغذى عكسياً كدخل للوحدة i خلال الوصلة w_{ij}، و α عامل التعليم. ينفَّذ تحديث الأوزان بعد استقرار الشبكة.

لاسترداد الشبكة، تحسب الوحدة رقم j الخرج تكرارياً بعد أن يقدم شعاع الدخل $\mathbf{x}$ للشبكة وفقاً للمعادلات (42.5) و(43.5) حتـــى تتوقف المخارج عن التغير. وكما في حالة شبكات هوبفيلد، يمكن أن يُعرَّف تابع الطاقة لشبكات BSB، ومن ثَمَّ يمكن تأكيد استقرار هذه الشبكات.

تصبح الشبكة مستقرة عندما تتوقف كل الوحدات عن التغير، وكما في حالة شبكات هوبفيلد يمكن أن يبقى تابع طاقة Lyapunov نفسه أو يتناقص بعد أن تتحدث الوحدات. توافق الطاقة الدنيا أوجه المكعب التـــي تمثل جواذب النقطة الثابتة. ويجري إثبات

الاستقرارية باتباع طريقة مشابهة لشبكات هوبفيلد المتقطعة أومن نظرية Cohen-Grossberg.

يمكننا تلخيص خوارزمية تدريب الشبكة وفقاً لـــ Anderson وزملائه عام 1977[47]، حيث تشبه هذه الخوارزمية ما اقترحه Hecht-Nielsonعام 1990 [46]، كما يلي:

1. ضع قيماً أولية للأوزان (قيم كيفية صغيرة)، ولمعدلات التدريب α وβ

2. لكل شعاع دخل تدريب كرِّر الخطوات 3 إلى 7.

3. ضع التفعيلات الأولية لوحدات الشبكة مساوية لشعاع الدخل الخارجي $\mathbf{x}$:

$$y_j = x_j$$

4. مادامت التفعيلات مستمرة في التغير نفذ الخطوات من 5 إلى 6.

5. حساب مداخل الشبكة

$$net_j = y_j + \alpha \sum_i y_i w_{ij}$$

(كل دخل للشبكة يتركب من التفعيل السابق والإشارات المستقبلة المثقلة من كل الوحدات)

6. كل وحدة تُعيِّن تفعيلها (إشارة خرجها) وفقاً لـــ:

$$y_j = \begin{cases} 1 & net_j > 1 \\ y_j & -1 \le net_j \le 1 \\ -1 & net_j < -1 \end{cases}$$

(الحالة المستقرة من أجل شعاع التفعيل ستكون أوجه المكعب)

7. تحديث الأوزان:

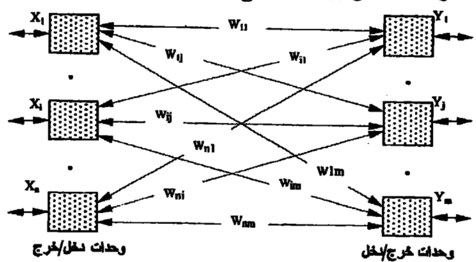
$$w_{ij}^{new} = w_{ij}^{old} + \beta\, y_i y_j$$

7.5 ذاكرة الترافق الثنائية الاتجاه

Bidirectional associative memory (BAM)

اقترح Kosko شبكات الذاكرة المترافقة الثنائية الاتجاه (BAM) عام 1987 [12]. وهذه الشبكات تكرارية وذات ترافق مغاير ومؤلفة من طبقتين، وهي تشبه إلى حد بعيد شبكات هوبفيلد وشبكات الطنين المتكيف (ART) (ستشرح فيما بعد) ماعدا أن كلا

الطبقتين تعملان وحدات دخل ووحدات خرج.



الشكل8.5: شبكة ذاكرة ترافق متغاير ثنائية الاتجاه

ستكون أشعة الدخل **x** ببعد n وأشعة الدخل **y** ببعد m، والوصلات بين الوحدات الثنائية الاتجاه، والنماذج يمكن أن تقدم للشبكة من أحد أطراف الوصلات الخارجية للشبكة كما هو موضح في الشكل (8.5).

تخزن مصفوفة الوزن **W** (لإرسال الإشارة من الطبقة X إلى الطبقة Y) ومصفوفة الوزن $\mathbf{W}^T$ (لإرسال الإشارة من الطبقة Y إلى الطبقة X) أزواج النماذج المترافقة. وهكذا يكون الوزن w_{ij} على الوصلة بين الوحدة i من طرف النموذج **x** إلى الوحدة j من طرف النموذج **y** مشتركاً لتدفق الإشارة في أي اتجاه. هذا وقد دُرِس التعليم المتكيف وغير المتكيف لشبكات BAM.

دُرِّبت الشبكة بسهولة باستعمال قاعدة Hebb، حيث تحسب مصفوفة الوزن كمجموع Pمصفوفة نموذج. أي:

$$\mathbf{W}^p = \mathbf{x}^p (\mathbf{y}^p)^T \qquad \mathbf{W} = \sum_{p=1}^{P} \mathbf{W}^p \qquad (46.5)$$

حيث

$$w_{ij} = \sum_{p=1}^{P} w_{ij}^p = \sum_{p=1}^{P} x_i^p y_j^p$$

لاسترداد نموذج، يقدم شعاع الدخل إما من الجانب الأيسر (الدخل **x**) أو من الجانب الأيمن (الدخل **y**) للشبكة، ويُحسب الدخل net_i للوحدة i كما يلي:

$$net_i = \sum_{j=1}^{n} w_{ij} x_j = \mathbf{W}_i . \mathbf{x} \qquad (47.5)$$

حيث $\mathbf{W}i$ هو السطر رقم i من مصفوفة الوزن **W**. يحسب الطرف المقابل (الطرف y) للشبكة تفعيلات الخرج وينتشر الناتج عكسياً إلى الطرف المقابل من خلال مصفوفة المنقول $\mathbf{W}^T$، تتكرر هذه العملية حتـــى تستقر الشبكة.

تعطى مصفوفة الأوزان لتخزين مجموعة من أشعة الدخل S(p) والخرج المنشود t(p) لــ P نموذج دخل، حيث:

$$s(p) = [s_1(p), s_2(p), \ldots, s_n(P)]$$

و

$$t(p) = [t_1(p), t_2(p), \ldots, t_n(P)]$$

في حالة أشعة دخل ثنائية كما يلي:

$$w_{ij} = \sum_{p} (2s_i(p) - 1)(2t_j(p) - 1) \qquad (48.5)$$

وفي حالة أشعة دخل ثنائية القطبية:

$$w_{ij} = \sum_{p} s_i(p) t_j(p) \qquad (49.5)$$

يحسب الخرج في اللحظة t + 1 باستخدام توابع التفعيل الثنائية القطبية للوحدات على الطرف y كما يلي:

$$y_j(t+1) = \mathrm{sgn}\left(\sum_{i=1}^{n} w_{ij} x_i(t)\right) = \begin{cases} +1 & w_{ij} x_i(t) > 0 \\ y_j & w_{ij} x_i(t) = 0 \\ -1 & w_{ij} x_i(t) < 0 \end{cases} \qquad (50.5)$$

لاحظ أن الوحدة رقم j تحتفظ بنفس القيمة عندما يكون الدخل التركيبـــي net مساوياً للصفر. يستخدم نفس النوع من توابع التفعيل للطرف x باستعمال المداخل التركيبية net من الطرف y. يمكن أحياناً استعمال قيمة عتبة بدلاً من الصفر وبذلك لا تتغير قيم الشبكة في مجال من القيم $[-\theta, +\theta]$. ويكون تابع التفعيل بعتبة $\theta > 0$:

$$y_j(t+1) = \mathrm{sgn}\left(\sum_{i=1}^{n} w_{ij} x_i(t)\right) = \begin{cases} +1 & w_{ij} x_i(t) > \theta \\ y_j & |w_{ij} x_i(t)| \le \theta \\ -1 & w_{ij} x_i(t) < -\theta \end{cases} \qquad (51.5)$$

بالطبع يمكن استخدام توابع تفعيل ثنائية أو Sigmoid لوحدات الشبكة.

على الرغم أن الحد الأعلى لسعة ذاكرة الشبكات العصبونية الصنعية المترافقة الثنائية الاتجاه هو $C < \min(n,m)$، حيث C السعة وهي العدد الأعظمي لأزواج النماذج التي تستطيع الشبكة تخزينها واستردادها بدقة، و n عدد وحدات الطبقة X و m عدد وحدات الطبقة Y (كما هو متوقع، فإن BAMs لا تستطيع تخزين أي نموذج زيادة على العدد الأعظمي للأسطر المستقلة خطياً (الأعمدة) في المصفوفة W)، فإن Haines و-Hechet Nielson عام 1988[48] أثبتا أن هذه السعة قد تُوسَّع لتكون $\min(2^n,2^m)$.

لقد وُجدت لشبكات الترافق المغاير الثنائية الاتجاه تطبيقات محدودة، ومع ذلك استعملت في معالجة الرؤية والتحكم (Kosko(a) عام 1988[42] وBavarian عام 1988[28]).

مثال 7:

حساب مسافة Hamming بين زوجين من الأشعة المترافقة،

يسمىّ عدد المركبات المختلفة في شعاعين x_1 وx_2 ممثلين بالشكل الثنائي أو ثنائي القطبية بمسافة هامنغ (Hamming) بين الشعاعين ويرمز لها بـــ $H[x_1,x_2]$. ومتوسط مسافة هامنغ بين الشعاعين هو $\frac{1}{n} H[x_1,x_2]$، حيث n هي عدد المركبات في كل شعاع.

فمثلاً، الأشعة x ببعد 3×5=15 مركبة (الممثلة بالحرفين A وC، وبالشكل الثنائي البعد) التالية:

0#0	0##
#0#	#00
###	#00
#0#	#00
#0#	0##
(-1,+1)	(+1,+1)

يختلف بعضها عن بعض في المركبات التالية: الثالثة والسادسة والثامنة والتاسعة والثانية عشرة والثالثة عشرة والرابعة عشرة. وهذا يعطي أن متوسط مسافة هامنغ بين هذين الشعاعين يساوي 7/15. ومتوسط مسافة هامنغ بين الشعاعين y ببعد 2 الممثلين بالشكل الثنائي القطبية المرافقين لشعاعي الدخل السابقين يساوي 1/2 (لاحظ أن المسافتان متقاربتين).

لقد لاحظ Kosko عام 1988[49] أن "ترميز الارتباط" (correlation encoding) كالذي استُعمل في الشبكة العصبونية الصنعية كذاكرة ترافق مغاير ثنائية الاتجاه سيتحسن عندما يكون متوسط مسافة هامنغ بين أزواج نماذج الدخل قابلاً للمقارنة (متقاربة نوعاً ما) مع متوسط مسافة هامنغ بين أزواج نماذج الخرج المرافقة. إذا تحققت هذه الحالة، ستطبق أو تحول مسافة هامنغ بين نماذج الدخل المفصولة بمسافة صغيرة إلى نماذج الخرج لتكون منفصلة بمسافة مشابهة، على حين أن أشعة الدخل المفصولة بعضها عن بعض بمسافة هامنغ كبيرة ستذهب لفصل نماذج الخرج على نحو كبير (غير مشابه للمسافة بين نماذج الدخل).

أما ما يخص حذف (erasing) الترافق المخزن فقد لوحظ أن متمم الشعاع الثنائي القطبية x يكوَّن بواسطة تغيير كل +1 بـــ -1 وبالعكس. واستنتج أن ترميز (encoding) (أي تخزين زوج النموذج) الزوج (s^c, t^c)، حيث s^c هو متمم الشعاع s، و t^c هو متمم الشعاع t، يخزن نفس المعلومات كترميز (s, t)، وترميز (s^c, t) أو (s, t^c) سيحذف ترميز (s, t)، هذا ما قاله Kosko(a) عام 1988[42].

1.7.5 تطبيقات شبكة الترافق الثنائية الاتجاه في التخزين والاستدعاء

مثال 8:

ليكن لدينا الزوجان التاليان من النماذج غير المتعامدة ممثلين بالشكل الثنائي القطبية:

$$(\mathbf{x}^1)^T = [+1\ -1\ +1\ -1\ +1\ -1]\ ,\ (\mathbf{y}^1)^T = [+1\ 1\ -1\ -1]$$

$$(\mathbf{x}^2)^T = [+1\ +1\ +1\ -1\ -1\ -1]\ ,\ (\mathbf{y}^2)^T = [+1\ -1\ +1\ -1]$$

باستعمال قاعدة Hebb، تحسب مصفوفات ارتباط النماذج لهذين الزوجين كما يلي:

$$\mathbf{W}^1 = \mathbf{x}^1(\mathbf{y}^1)^T = \begin{bmatrix} +1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 \\ +1 & +1 & -1 & -1 \\ -1 & -1 & +1 & +1 \end{bmatrix}$$

$$W^2 = x^2(y^2)^T = \begin{bmatrix} +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ -1 & +1 & -1 & +1 \\ -1 & +1 & -1 & +1 \\ -1 & +1 & -1 & +1 \end{bmatrix}$$

مصفوفة الوزن **W** اللازمة لتخزين كلا زوجي النموذجين ستكون:

$$W = W^1 + W^2 = \begin{bmatrix} +2 & 0 & 0 & -2 \\ 0 & -2 & +2 & 0 \\ 2 & 0 & 0 & -2 \\ -2 & 0 & 0 & +2 \\ 0 & +2 & -2 & 0 \\ -2 & 0 & 0 & +2 \end{bmatrix}$$

سنختبر مقدرة الشبكة على استرداد النموذج **y** المرافق للشعاع **x** المطبق على الطبقة **X** كدخل للشبكة:

$$(x^1)^T W = [+1 -1 +1 -1 +1 -1] \begin{bmatrix} +2 & 0 & 0 & -2 \\ 0 & -2 & +2 & 0 \\ +2 & 0 & 0 & -2 \\ -2 & 0 & 0 & +2 \\ 0 & +2 & -2 & 0 \\ -2 & 0 & 0 & +2 \end{bmatrix}$$

$$= [+8 \ +4 \ -4 \ -8]$$

$$f([+8 \ +4 \ -4 \ -8]) = [+1 \ +1 \ -1 \ -1] = (y^1)^T$$

وسنختبر مقدرة الشبكة على استرداد النموذج $\mathbf{x}^1$ المرافق للشعاع $\mathbf{y}^1$ المطبق على الطبقة **Y** كدخل للشبكة:

$$(\mathbf{y}^1)^T \mathbf{W}^T = \mathbf{W}\mathbf{y}^1 = \begin{bmatrix} +2 & 0 & 0 & -2 \\ 0 & -2 & +2 & 0 \\ +2 & 0 & 0 & -2 \\ -2 & 0 & 0 & +2 \\ 0 & +2 & -2 & 0 \\ -2 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} +1 \\ +1 \\ -1 \\ -1 \end{bmatrix} = \begin{bmatrix} +4 \\ -4 \\ +4 \\ -4 \\ +4 \\ -4 \end{bmatrix}$$

$$f\left(\begin{bmatrix} +4 \\ -4 \\ +4 \\ -4 \\ +4 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} = \mathbf{x}^1$$

سنحاول الآن تطبيق دخل ضجيجي $\mathbf{x}^*$ (على الطرف X) المكوَّن من الشعاع $\mathbf{x}^2$ بوضع المركبة الأول خطأ كما يلي:

$$(\mathbf{x}^*)^T = [-1 +1 +1 -1 -1 -1]$$

وسنحاول استرداد النموذج المرافق:

$$(\mathbf{x}^*)^T \mathbf{W} = [-1 +1 +1 -1 -1 -1] \begin{bmatrix} +2 & 0 & 0 & -2 \\ 0 & -2 & +2 & 0 \\ +2 & 0 & 0 & -2 \\ -2 & 0 & 0 & +2 \\ 0 & +2 & -2 & 0 \\ -2 & 0 & 0 & +2 \end{bmatrix}$$

$$= [+4 -4 +4 -4]$$

$$f([+4 -4 +4 -4]) = [+1 -1 +1 -1] = (\mathbf{y}^*)^T = \mathbf{y}^2$$

ومن الطرف Y للشبكة نجد ما يلي:

$$(\mathbf{y}^2)^T \mathbf{W}^T = \mathbf{W}\mathbf{y}^2 = \begin{bmatrix} +2 & 0 & 0 & -2 \\ 0 & -2 & +2 & 0 \\ +2 & 0 & 0 & -2 \\ -2 & 0 & 0 & +2 \\ 0 & +2 & -2 & 0 \\ -2 & 0 & 0 & +2 \end{bmatrix} \begin{bmatrix} +1 \\ -1 \\ +1 \\ -1 \end{bmatrix} = \begin{bmatrix} +4 \\ +4 \\ +4 \\ -4 \\ -4 \\ -4 \end{bmatrix}$$

$$f\left(\begin{bmatrix} +4 \\ +4 \\ +4 \\ -4 \\ -4 \\ -4 \end{bmatrix}\right) = \begin{bmatrix} +1 \\ +1 \\ +1 \\ -1 \\ -1 \\ -1 \end{bmatrix} = \mathbf{x}^2$$

مثال 9:

شبكة ترافق ثنائية الاتجاه لتخزين أحرف مع رموز ثنائية القطبية مترافقة،

المطلوب تصميم شبكة BAM متقطعة مؤلفة من 15 وحدة دخل ووحدتــي خرج لتخزين زوجين من النماذج المترافقة؛ الزوج الأول الحرفان A وC (معطاة بـــ 5× 3= 15 مركبة) والزوج الثانــي رمز ثنائي القطبية مؤلف من مركبتين، كما هو مبين.

تُمثّل الأحرف بالبعد الثنائي بالرمز "#" إذا كانت النقطة مضاءة من الحرف (1+) وبالرمز "0" إذا لم تكن النقطة مضاءة من الحرف (−1).

0#0	0##
#0#	#00
###	#00
#0#	#00
#0#	0##
(-1,1)	**(1,1)**

ستكون مصفوفات الأوزان لتخزين النموذج A والنموذج B وكلا النموذجين كما يلي:

W لتخزين كلا النموذجين $\qquad$ لتخزين A ← (−1,1) $\qquad$ لتخزين C ← (1,1)

$$\begin{bmatrix} 0 & -2 \\ 0 & +2 \\ +2 & 0 \\ 0 & +2 \\ 0 & -2 \\ -2 & 0 \\ 0 & +2 \\ -2 & 0 \\ -2 & 0 \\ 0 & +2 \\ 0 & -2 \\ -2 & 0 \\ -2 & 0 \\ +2 & 0 \\ 0 & +2 \end{bmatrix} \qquad \begin{bmatrix} +1 & -1 \\ -1 & +1 \\ +1 & -1 \\ -1 & +1 \\ +1 & -1 \\ -1 & +1 \\ -1 & +1 \\ -1 & +1 \\ -1 & +1 \\ -1 & +1 \\ +1 & -1 \\ -1 & +1 \\ -1 & +1 \\ +1 & -1 \\ -1 & +1 \end{bmatrix} \qquad \begin{bmatrix} -1 & -1 \\ +1 & +1 \\ +1 & +1 \\ +1 & +1 \\ -1 & -1 \\ -1 & -1 \\ +1 & +1 \\ -1 & -1 \\ -1 & -1 \\ +1 & +1 \\ -1 & -1 \\ -1 & -1 \\ -1 & -1 \\ +1 & +1 \\ +1 & +1 \end{bmatrix}$$

بالطبع علينا أولاً تحويل الأحرف إلى أشعة ثنائية القطبية كما فعلنا من قبل في كثير من الأمثلة السابقة، وسنهمل هذه العملية هنا بغية الاختصار. أما مصفوفات الوزن **W** اللازمة لتخزين النماذج المترافقة فهي مبينة آنفاً.

لتوضيح استعمال شبكة الترافق الثنائية الاتجاه سنرى استجابة الشبكة عندما تعطى الشعاع **x** كدخل (إما الحرف A أو C):

نموذج الدخل A:

$$[-1\ 1\ -1\ 1 -1\ 1\ 1\ 1\ 1\ 1\ -1\ 1\ 1\ -1\ 1]\cdot \mathbf{W} = (-14,16) \xrightarrow{f(net)} (-1,1)$$

نموذج الدخل C:

$$[-1\ 1\ 1\ 1 -1\ -1\ 1\ -1\ -1\ 1\ -1 -1 -1\ 1\ 1]\cdot \mathbf{W} = (14,18) \xrightarrow{f(net)} (1,1)$$

ولتوضيح ثنائية اتجاه الشبكة سنقدم الأشعة **y** كدخل. لإرسال الإشارة من الطبقة Y إلى الطبقة X، ستكون مصفوفةُ الوزن منقولَ المصفوفة **W**:

$$W^T = \begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

لاسترداد النموذج A المرافق للرمز(1,1−) سنطبق هذا الرمز على الطبقة Y كدخل:

$$W= (-1,1) . (-1,1)$$

$$\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$=[-2\ 2\ -2\ 2\ -2\ 2\ 2\ 2\ 2\ 2\ -2\ 2\ 2\ -2\ 2]$$

$$\xrightarrow{f(net)} [-1\ 1\ -1\ 1\ -1\ 1\ 1\ 1\ 1\ 1\ -1\ 1\ 1\ -1\ 1]$$

وهو النموذج A كشعاع ثنائي القطبية. وبالمثل يجري استرداد النموذج C بتطبيق الرمز (1,1) كدخل على الطبقة Y:

$$W = (1,1).$$

$$(1,1).\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$= [-2\ 2\ 2\ 2\ -2\ -2\ 2\ -2\ -2\ 2\ -2\ -2\ -2\ 2\ 2]$$

$$\xrightarrow{f(net)} [-1\ 1\ 1\ 1\ -1\ -1\ 1\ -1\ -1\ 1\ -1\ -1\ -1\ 1\ 1]$$

وهو النموذج C.

ويمكن استخدام هذه الشبكة بمداخل ضجيجية لـــ x وy أو كليهما على النحو التالي: سنطبق على دخل الشبكة y (الطبقة Yكمدخل) نسخة ضجيجية (0,1) عن الرمز المخزن (−1,+1) وليس هناك معلومات عن شعاع x الموافق (هذا يعنـــي أن الشعاع x سيكون مطابقاً للصفر)، والضجيج هو ضياع في المعطيات.

$$(0,1).\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$

$$= [-2\ 2\ 0\ 2\ -2\ 0\ 2\ 0\ 0\ 2\ -2\ 0\ 0\ 0\ 2]$$

$$\xrightarrow{f(net)} [-1\ 1\ 0\ 1\ -1\ 0\ 1\ 0\ 0\ 1\ -1\ 0\ 0\ 0\ 1]$$

لاحظ أن الوحدات التـــي لها net يساوي الصفر يبقى تفعيلها ثابتاً. هذا الشعاع سيرسَل

إلى الخلف إلى الطبقة Y باستعمال المصفوفة W:

$$[-1\ 1\ 0\ 1\ -1\ 0\ 1\ 0\ 0\ 1\ -1\ 0\ 0\ 0\ 1]\begin{bmatrix} 0 & -2 \\ 0 & +2 \\ +2 & 0 \\ 0 & +2 \\ 0 & -2 \\ -2 & 0 \\ 0 & +2 \\ -2 & 0 \\ -2 & 0 \\ 0 & +2 \\ 0 & -2 \\ -2 & 0 \\ -2 & 0 \\ +2 & 0 \\ 0 & +2 \end{bmatrix}$$

$$=[0\quad 16]\xrightarrow{f(net)}(0,1)$$

هذه النتيجة ليست غريبة لأن الشبكة لا تملك أي معلومات مرجعية لـــ A أو C. وقد تقاربت الشبكة (لأن من الواضح أنه لن تحدث تغيرات في التفعيلات) إلى حالة استقرار وهمية؛ وهذا يعنـــي أن الحل ليس واحداً لأزواج النماذج المخزنة.

من ناحية أخرى، إذا أُعطيت الشبكة شعاعَ الدخل **y** السابق وبعض المعلومات حول الشعاع **x** مثلاً،

$$\mathbf{y}=(0,1),\quad \mathbf{x}=[0\ 0\ -1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ -1\ 0]$$

فإن الشبكة ستكون قادرة على الوصول إلى الحالة المستقرة والاستجابة الصحيحة الموافقة لأزواج النماذج المخزنة. لاحظ أن الشعاع x نسخة ضجيجية من A:

$$A=[-1\ 1\ -1\ 1\ -1\ 1\ 1\ 1\ 1\ 1\ -1\ 1\ 1\ -1\ 1]$$

والمركبات التـــي لا تساوي الصفر في **x** هي التـــي تميز A عن C:

$$C = [-1\ 1\ 1\ 1\ -1\ -1\ 1\ -1\ -1\ 1\ -1\ -1\ -1\ 1\ 1]$$

وباعتبار أن خوارزمية التدريب تبقي خرج الوحدة ثابتاً عندما يكون دخل الوحدة net مساوياً للصفر فإن:

$$W^T = (0,1)\cdot(0,1)\cdot\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$
$$= [-2\,2\,0\,2\,-2\,0\,2\,0\,0\,2\,-2\,0\,0\,0\,2]$$

$$\xrightarrow{f(net)} [-1\ 1\ -1\,1\ -1\,1\,1\,1\,1\,1\ -1\,1\,1\ -1\,1]$$

وهو النموذج A. لاحظ أننا في هذه الحالة أفرطنا في إعطاء الشبكة معلومات عن x (كل مركبة تميز A عن C أعطيت قيمة دخل لـــ A). دعنا الآن نفعل الشيء نفسه ولكن بمعلومات أقل معطاة عن الشعاع x. مثلاً، ليكن

$$y = (0,1) \text{ و } x = [0\ 0\ -1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 0\ 0\ 0]$$

عندئذ ستكون استجابة الشبكة:

$$W^T = (0,1)\cdot(0,1)\cdot\begin{bmatrix} 0 & 0 & 2 & 0 & 0 & -2 & 0 & -2 & -2 & 0 & 0 & -2 & -2 & 2 & 0 \\ -2 & 2 & 0 & 2 & -2 & 0 & 2 & 0 & 0 & 2 & -2 & 0 & 0 & 0 & 2 \end{bmatrix}$$
$$= [-2\,2\,0\,2\,-2\,0\,2\,0\,0\,2\,-2\,0\,0\,0\,2]$$
$$\xrightarrow{f(net)} [-1\ 1\ -1\,1\ -1\,1\,1\,1\,0\,1\ -1\,0\,0\,0\,1]$$

وهذا الشعاع ليس تماماً النموذج A. لذا سنلجأ إلى العملية التكرارية وذلك بإرسال الشعاع الناتج إلى الخلف إلى الطبقة Y باستعمال مصفوفة الوزن W:

وإذا غذي هذا النموذج للخلف إلى الطبقة X مرة أخرى ستعطي الشبكة النموذج A تماماً.

$$
\begin{bmatrix}-1 & 1 & -1 & 1 & -1 & 1 & 1 & 1 & 0 & 1 & -1 & 0 & 0 & 0 & 1\end{bmatrix}\begin{bmatrix}0 & -2\\ 0 & +2\\ +2 & 0\\ 0 & +2\\ 0 & -2\\ -2 & 0\\ 0 & +2\\ -2 & 0\\ -2 & 0\\ 0 & +2\\ 0 & -2\\ -2 & 0\\ -2 & 0\\ +2 & 0\\ 0 & +2\end{bmatrix}
$$

$$
=[-6\,16]\xrightarrow{f(net)}(-1,1)
$$

8.5 تمارين

1.5 استعمل قاعدة Hebb لتخزين الأشعة (1, 1, 1, 1) و(1- ,1- ,1 ,1) في شبكة عصبونية مترافقة ذاتياً.

1. أوجد مصفوفة الوزن (لا تضع عناصر القطر الرئيسي أصفاراً).
2. اختبر الشبكة باستعمال الشعاع (1, 1, 1, 1)كدخل.
3. اختبر الشبكة باستعمال الشعاع (1- ,1- ,1 ,1) كدخل.
4. اختبر الشبكة باستعمال (0 ,1 ,1 ,1) كدخل وناقش النتائج.
5. كرر الطلبات من 4.1 بعناصر صفرية على القطر الرئيسي في مصفوفة الأوزان. ناقش الاختلاف في استجابة الشبكة لكلا الحالتين.

2.5 ليكن لدينا شبكة ترافق ذاتـــي بتابع خطوة ثنائي القطبية والأوزان معينة بقاعدة Hebb (الجداءات الخارجية)، وبعناصر صفرية على القطر الرئيسي في مصفوفة الأوزان.

1. أوجد مصفوفة الوزن لتخزين الشعاع $V_1=(1, 1, 1, 1, 1, 1)$

2. اختبر الشبكة باستعمال V_1 كدخل

3. اختبر الشبكة باستعمال $T_1 = (1, 1, 1, 1, -1, -1)$

4. أوجد مصفوفة الوزن لتخزين الشعاع $V_2 = (1, 1, 1, -1, -1, -1)$

5. اختبر الشبكة باستعمال V_2 كدخل

6. اختبر الشبكة باستعمال $T_2 = (1, 1, 1, -1, 0, 0)$

7. أوجد مصفوفة الوزن لتخزين كلٍّ من V_2, V_1

8. اختبر الشبكة على V_1, V_2, T_1, T_2

3.5 باستعمال قاعدة تدريب Hebb لشبكة ترافق ثنائية الاتجاه

1. أوجد مصفوفة الأوزان الثنائية القطبية لتخزين أزواج نماذج الدخل /الخرج الثنائية التالية:

$$x(1) = (1, 0, 1) \quad y(1) = (1, 0)$$

$$x(2) = (0, 1, 0) \quad y(2) = (0, 1)$$

2. استعمل تابع الخطوة الثنائي (بعتبة θ)كتابع تفعيل لجميع وحدات الطبقتين، اختبر استجابة شبكتك في كلا الاتجاهين على كل نماذج تدريب الدخل.اعتبر تفعيلات طبقة الخرج تساوي الصفر عندما تطبق نموذج تدريب على الدخل.

3. استعمل تابع تفعيل ثنائي القطبية (بعتبة θ) كتابع تفعيل لوحدات كلا الطبقتين،حوِّل نماذج التدريب إلى الشكل الثنائي القطبية واختبر استجابة الشبكة في كلا الاتجاهين ثانية. ضع التفعيلات الأولية كما في الطلب السابق.

4. اختبر استجابة الشبكة على كل نسخ الضجيج التالية الثنائية القطبية، كرر الخوارزمية حتـــى تستقر الشبكة:

$$a. (0, -1, 1) \quad b. (0, 0, 1) \quad c. (1, 0, 0)$$

$$d. (-1, 0, -1) \quad e. (-1, 0, 0) \quad f. (0, 0, -1)$$

5. في أية حالة تستقر الشبكة على الاستجابة الصحيحة، وأية حالة تكون الاستجابة عندها غير معروفة أو غير محددة؟

4.5 استعمل الجداء الخارجي لقاعدة تعليم Hebb

1. لإيجاد مصفوفة الوزن بالشكل الثنائي من أجل شبكة ذاكرة ترافق ثنائية الاتجاه لتخزين

أزواج أشعة الدخل/الخرج الثنائية:

s (1) = (1, 0, 0, 0)　　　t (1) = (1, 0)

s (2) = (1, 0, 0, 1)　　　t (2) = (1, 0)

s (3) = (0, 1, 0, 0)　　　t (3) = (0, 1)

s (4) = (0, 1, 1, 0)　　　t (4) = (0, 1)

2. استعمل تابع الخطوة (مع عتبة تساوي 0) كتابع تفعيل وحدات الخرج، اختبر الشبكة على كل نماذج الدخل. ناقش جميع النتائج.

3. اختبر استجابة الشبكة على تراكيب متنوعة من نماذج الدخل مع " أخطاء" في المعطيات أو "ضياع" في المعطيات، ناقش النتائج التـــي حصلت عليها.

5.5 اكتب برنامجاً لأداء شبكة هوبفيلد المتقطعة وذلك لتخزين الأحرف من أحد التشكيلات في الشكل (31.4) .المثال 16 (في الفقرة 4.5.4 من الفصل الرابع). حقق في عدد النماذج التـــي يمكن أن تخزن وتستدعى على نحو صحيح، الإضافة إلى إمكانية الشبكة للا[illegible] لدخل ضجيجي.

6.5 اكتب برنامجاً لأداء شبكة عصبونية كذاكرة ترافق مغاير ثنائية الاتجاه.تتألف على اذفر من 15 وحدة في الطبقة X وثلاثة وحدات في الطبقةY.

1. استعمل البرنامج لتخزين الترافق المعطى في المثال 9 السابق؛ أي:

```
0#0        0##
#0#        #00
###        #00
#0#        #00
#0#        0##
(-1,1)     (1,1)
```

2. حاول توضيح نفس الحالات المناقشة في المثال، وبعض الحالات الخاصة بك. اختبر الشبكة في حالة دخل ضجيجي للحرف C مثلاً.

3. استعمل برنامجك لتخزين النماذج التالية (أشعة الطبقة X تعطى بمصفوفات 3×5 وأشعة الطبقة Y معطاة كما يلي لكل نموذج X:

4. هل يمكن تخزين كل النماذج الثمانية السابقة؟ إذا كان لا، كم عدد النماذج التـــي يمكن تخزينها معاً؟ حاول ببعض التجارب الحصول على نسخ ضجيجية كما في الطلب الأول.

5. الجدول التالي يعطيك مسافة هامنغ بين النماذج السابقة المشار لها بالأحرف المشابهة:

	A	B	C	D	E	F	G	H
A	0	4	7	4	6	6	5	3
B		0	7	2	4	4	7	5
C			0	7	3	5	2	8
D				0	6	6	5	5
E					0	2	5	5
F						0	8	5
G							0	6
H								0

حدد مسافة هامنغ بين نماذج الطبقة-Y المترافقة مع كل هذه الأحرف. من نسب مسافات هامنغ، حاول تعيين أي أزواج النماذج سيكون مخزناً بنجاح أكثر.

6. باعتبار أن الحد الأعلى لعدد أزواج النماذج الكيفية التـــي يمكن أن تخزن هو min(n,m)، اختبر بدقة أية حالة يكون فيها عدد النماذج المخزنة أكثر. اختبر استجابة الشبكة في الاتجاهين.

7.5 اكتب برنامجاً لأداء شبكة عصبونية كذاكرة ترافق ذاتـــي باستعمال قاعدة Hebb لتشكيل مصفوفة الأوزان (بالجداءات الخارجية). يقرأ البرنامج الدخل من مصفوفة 5×7 إلى الشعاع x (35 مركبة) وسيكون لبرنامجك حالة تدريب يقوم فيها بحساب قيم مصفوفة الأوزان (بالجداءات الخارجية).

يمكن تحديد عدد المداخل في بداية البرنامج، ولكن يجب أن يكون كافياً لمحاولة إدخال نماذج أكبر. وسيكون للبرنامج أيضاً حالة اختبار تكون فيه الأوزان ثابتة واستجابة الشبكة محددة. الاستجابة ستطبع على الخرج كمصفوفة 5×7. ومن المفيد عرض مصفوفة الدخل أيضاً متبوعة بمصفوفة الخرج للمقارنة. نماذج الدخل الواجب تخزينها هي:

```
00.00   0...0   0...0   0000.   .....   0....   .....   0...0   0...0   0...0
0..00   .000.   .000.   000..   .0000   .0000   0000.   .000.   .000.   .000.
00.00   0000.   0000.   00.0.   .0000   .0000   000.0   .000.   .000.   .000.
00.00   000.0   00..0   0.00.   ....0   ....0   000.0   0...0   0....   .000.
00.00   00.00   0000.   .....   0000.   .000.   00.00   .000.   0000.   .000.
00.00   0.000   .000.   0000.   0000.   .000.   00.00   .000.   0000.   .000.
0...0   .....   0...0   0000.   ....0   0...0   0.000   0...0   ....0   0...0
```

حاول الإجابة على الأسئلة التالية:

1. ماهو عدد النماذج التـــي يمكن أن تخزن وتستدعى على نحو صحيح؟

2. ماهو مقدار الضجيج الذي يمكن أن تعالجه شبكتك؟

8.5 اكتب برنامجاً لأداء شبكة عصبونية كذاكرة ترافق مغاير باستعمال قاعدة Hebb لتشكيل مصفوفة الأوزان (بواسطة الجداءات الخارجية). سيقرأ البرنامج لشعاع x من مصفوفة 7×9 والشعاع y من مصفوفة 5×3.

ابدأ باستعمال النماذج في المثال 3 من الفقرة 2.3.5 (الأشكال (4.5) و(6.5). وسّع خوارزميتك لتشمل أحرفاً أكثر، اعتمد النماذج x من تشكيلة واحدة من التشكيلات في الشكل (31.4) المثال 16 (في الفقرة 4.5.4 من الفصل الرابع)، أو اصنع تشكيلات خاصة بك، واستعمل نماذج الخرج في التمرين (6.5) السابق، أو اعتمد أشعة خاصة بك. ما هو عدد أزواج النماذج التـــي يمكن أن تخزن في شبكتك؟ اختبر استجابة الشبكة لدخل ضجيجي.

يمكن أن تجد مـــن المناسب تمثيل نماذج تدريبك بمصفوفات بإدخال "2" إذا كانت النقطة (pixel) on و"0" إذا كانت off. بعدئذ سيطرح برنامجك "واحد" من كل إدخال ليكوِّن شعاع نموذج ثنائي القطبية. هذا سيجعلك تمثل المعطيات الضائعة (من أجل أشعة الاختبار) بـــ "1" وسيقوم برنامجك بقلبها إلى الصفر متـــى رغبت بذلك.

الفصل السادس

الشبكات العصبونية المتعددة الطبقات بتغذية أمامية والانتشار الخلفي

Multilayer Feedforward Neural Networks and Backpropagation

سنصف في هذا الفصل واحدة من أهم بنـــى الشبكات العصبونية الصنعية الشائعة، إنها الشبكات ذات التغذية الأمامية، المتعددة الطبقات MLFF التـــي دربت باستعمال خوارزمية الانتشار الخلفي Bp (Backpropagation). يسمى هذا النوع من الشبكات أحياناً بيرسبترون متعدد الطبقات بسبب التشابه الكبير بين هذه الشبكات وشبكات البيرسبترون بأكثر من طبقة واحدة. سنبدأ بوصف بنية الشبكة وخوارزمية تعليم الانتشار الخلفي Bp وسننتهي إلى تعليم قاعدة دلتا المُعمَّمة وكيفية تنفيذ هذه الشبكات عملياً.

1.6 تمهيد

افترض أننا نرغب بتدريب شبكة عصبونية لتنفيذ تطبيق (mapping) كيفي $g': \mathbf{x} \rightarrow \mathbf{z}$، حيث $\mathbf{X}$ و$\mathbf{Z}$ أشعة عشوائية في $\mathbf{R}^n$ و $\mathbf{R}^m$ على الترتيب بتوزيع احتمالي مشترك $\Psi(\mathbf{x},\mathbf{z})$. إن ما هو متوفر لدينا لتدريب هذه الشبكة هو مجموعة التدريب المؤلفة من P زوج تدريب $\{\mathbf{x}^p,\mathbf{t}^p \mid p=1,2,\ldots,P\}$، ويتوفر لنماذج الدخل $\mathbf{X}^p$ توزيعاً احتمالياً معروفاً $\rho(\mathbf{x})$، و$\mathbf{t}^p$ هي قيم شعاع الخرج المنشود المعروفة لأشعة الدخل الموافقة $\mathbf{x}^p$، والخرج المحسوب بواسطة الشبكة هو $\mathbf{z}^p$ في حالة دخل معطى $\mathbf{x}^p$، حيث $\mathbf{z} = \mathbf{g}(\mathbf{x})$.

لقد رأينا فيما سبق من الفصول أن الشبكات العصبونية وحيدة الطبقة (مثل البيرسبترون

البسيط وأدلين و.....) تعوزها المقدرة على حساب توابع كيفية معقدة كتلك الحدود غير الخطية الفاصلة بين مناطق صفوف أشكال ونماذج الدخل المختلفة. ويعتبر هذا واحداً من أهم القيود الخطيرة للشبكات الأساسية بدون طبقات داخلية.

تعتبر الأنظمة الخطية متعددة الطبقات (مثل الشبكة العصبونية بطبقتين مع عصبونين McCulloch-Pitts في الطبقة المخفية المصممة لتنفيذ التابع XOR التـــي جرى شرحها في الفصل الرابع) مكافئة للنظام الخطي بطبقة واحدة ، ومن ثمَ فإن هذه الأنظمة أيضاً تفتقر إلى المقدرة على تنفيذ التطبيقات المعقدة الهامة.

سنصمم في هذا الفصل شبكات تغذية أمامية لها على الأقل طبقة واحدة من العصبونات، حيث سيكون كل عصبون قادراً على حساب تابع تفعيل غير خطي. وسنناقش بعض الشبكات المتعددة الطبقات بتغذية أمامية، وسنرى أن مثل هذه الشبكات يمكن أن تُدرَّب لحساب توابع غير خطية كيفية.

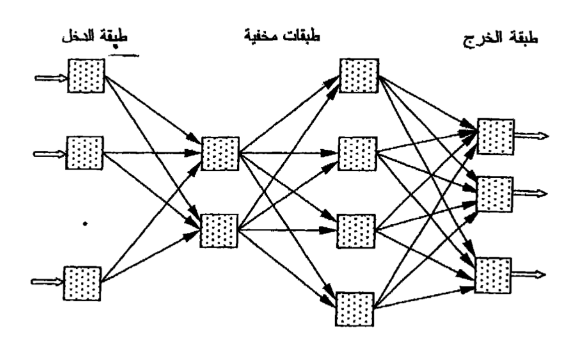
2.6 الشبكات متعددة الطبقات بتغذية أمامية

Multilayer feedforward networks (MLFF)

يوضح الشكل (1.6) بنية الشبكة المتعددة الطبقات بتغذية أمامية MLFF، حيث تتصل مكونات (أو عناصر) هذه الشبكة اتصالاً كاملاً وبتسلسل تصاعدي. تتألف الشبكة من طبقة دخل، وطبقة واحدة أو أكثر من الطبقات الداخلية المخفية، وطبقة خرج.

تسمى الطبقات الداخلية بالطبقات المخفية لأنها تستقبل مداخل داخلية من وحدات معالجة أخرى، وتعطي مخارج داخلية أيضاً إلى وحدات معالجة أخرى، ومن ثمَ فهي مخفية عن المستخدم الذي لا يتعامل معها مباشرة.

تعطى أشعة الدخل **x** ببعد n (ذات القيم الحقيقية) إلى كل وحدات الطبقة المخفية الأولى من خلال الأوزان w_{ij}. فتستلم وحدة الطبقة المخفية **j** الدخل **i** من خلال الوزن w_{ij}، حيث $i=1,2,...,n$ و $j=1,2,...,h$، ومن ثم تقوم بحساب تابع إشارة مداخلها netj من شعاع الدخل **x** والأوزان w_{ij}، وتمرر خرجها إلى الأمام إلى كل الوحدات في الطبقة التالية مباشرة.



الشكل 1.6: شبكة متعددة الطبقات بتغذية أمامية عامة.

وبالمثل، تتصل جميع وحدات الطبقة المخفية الثانية مع الطبقة السابقة لها من خلال الأوزان، وتقوم هذه الوحدات أيضاً بحساب تابع مداخلها net وتمرير خرجها إلى الطبقة التالية. تتكرر هذه العملية حتـــى ينجز الحساب النهائي بواسطة وحدات الخرج.

لقد أجريت عدة تعديلات على الشبكة MLFF الأساسية ودرست عبر السنين التالية لتصميم البيرسبترون من قبل Rosenblatt عام 1958[21] وعام 1961[50]. وقد أصبحت بعض الأنظمة منفذة عملياً كوحدات تعرّف الأشكال بعد إيجاد خوارزمية تدريب الانتشار الخلفي Bp التـــي سنعرضها فيما بعد.

اكتسبت الشبكات العصبونية الصنعية المتعددة الطبقات شعبيتها بسبب قدرتها على إنجاز تطبيقات كيفية $g:R^n \to R^m$، حيث $g(x) = z$. يمكن تنفيذ بعض التطبيقات إذا كان هناك عددٌ كافٍ من الوحدات المخفية وجرى تدريب الشبكة؛ أي وُجدت مجموعة من الأوزان التـــي تنجز مخططاً مرغوباً به.

لقد رأينا فـــي الفصول السابقة أن طريقة تدريب متوسط المربعات الصغرى لـــ Widrow-Hoff أو قاعدة تعليم دلتا مشابهة لخوارزمية تعليم البيرسبترون ماعدا أنها مُعرَّفة من أجل توابع تفعيل خطية.

وباستعمال تابع تفعيل تفاضلـــي نصبح قادرين على تعريف تابع خطأ (أو تابع الكلفة

cost) تفاضلي يمكن منه أن تتطور طريقة تعليم تدرج الهبوط. تلك القاعدة التـــي لها الشكل التالي:

$$w_{ij} = \eta \delta_j x_{ij} \qquad (1.6)$$

نُذكِّر أن هذه القاعدة ستكون مثلى عندما توجد مجموعة من الأوزان تجعل متوسط مربع الخطأ أصغرياً.

لسوء الحظ، لا يمكن أن تطبق خوارزمية تعليم البيرسبترون ولا قاعدة دلتا لـــ Widrow-Hoff على الشبكات المتعددة الطبقات؛ فقد أخفقت جميع هذه الطرق في معرفة طريقة تعديل أوزان الطبقات المخفية. وقد عُرِف هذا بمشكلة تعيين ـــ الاعتماد (credit-assignment). لأن هذه الطرق غير قادرة على تقدير الاعتماد أو تحديد مسؤولية أوزان الطبقة المخفية في تخفيض الأخطاء التـــي تحدث في طبقة الخرج.

لقد استطاعت طريقة تعليم الانتشار الخلفي، التـــي هي تعميم لقاعدة دلتا، تحديد كيفية تعديل الأوزان في الطبقات المخفية، ومن ثَمَّ سمحت لنا بإنشاء شبكات متعددة الطبقات بتغذية أمامية يمكنها تعلم مخططات **g** من $\mathbf{R}^n$ إلى $\mathbf{R}^m$ أكثر تعقيداً.

3.6 قاعدة دلتا المعممة أو خوارزمية تعليم الانتشار الخلفي

اكتُشفت طريقة تعليم الانتشار الخلفي من قبل الكثير من الباحثين على نحو مستقل وفي أزمنة مختلفة ولأسباب مختلفة، بدءاً من Werbos عام 1974[51] الذي اقترح استعمالها في أطروحته للدكتوراه، بعنوان "التراجع الخلفي: أدوات من أجل التنبؤ والتحليل في العلوم السلوكية"، بجامعة Harvard. ومروراً بـــ Parker عام 1985[52] الذي اكتشفها في تقريره العلمي الذي أعده في معهد ماستشوست للتقانه MIT على "منطق التعليم". ومع ذلك فإن الفضل ينسب عادة لـــ Rumelhart وزملائه في مجموعة المعالجة الموزعة التفرعية عام [53] 1985 لنشرهم الخوارزمية في المقالات العلمية وتطويرها لتتحول إلى إجراء قابل للعمل.

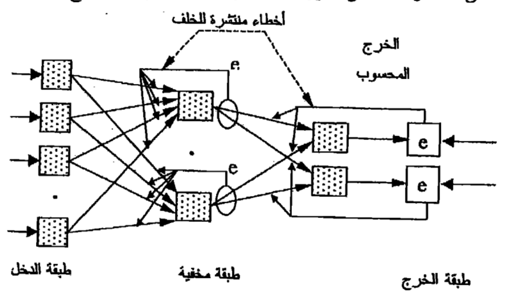
يمكن أن تطبق هذه الطريقة على أية شبكة متعددة الطبقات تستعمل توابع تفعيل تفاضلية والتعليم بمعلم. وكما في قاعدة دلتا، فهي إجراء أمثلية أُسِّسَ على تدرج الهبوط، حيث تعدل الأوزان لتقليل خطأ النظام أو تابع الكلفة. أتـــى اسم الانتشار الخلفي من الطريقة التـــي

يجري وفقها وضع الأوزان على الارتباطات.

تقدَّم نماذج الدخل إلى الشبكة، خلال طور التعليم، في شكل متتالية ما. ينتشر كل نموذج إلى الأمام طبقة بعد أخرى حتـــى يتم حساب نموذج الخرج. بعدئذ يقارن الخرج المحسوب مع الخرج المنشود لتحديد قيمة الخطأ.

تستعمل هذه الأخطاء مداخلَ لوصلات التغذية العكسية لتُعدَّل الأوزان طبقة بعد أخرى في الاتجاه العكسي حتـــى الوصول إلى أول طبقة مخفية من طرف الدخل. يوضح الشكل (2.6) بنية شبكة متعددة الطبقات بتغذية أمامية للانتشار الخلفي.

تُستعمل الوصلات العكسية خلال طور التعليم فقط، أما الوصلات الأمامية فتستعمل خلال طور التعليم والأطوار العملية الأخرى. ففي خوارزمية الانتشار الخلفي، تعدَّل أوزان الطبقة المخفية باستعمال الأخطاء من الطبقة التالية، وهكذا حتـــى الوصول إلى طبقة الخرج حيث تُستعمل أخطاؤها لتعديل الأوزان بين آخر طبقة مخفية وطبقة الخرج.



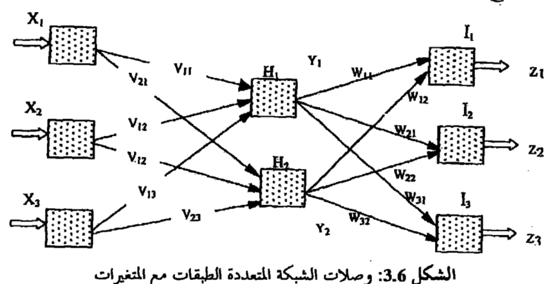
الشكل 2.6: شبكة تغذية أمامية بطبقة مخفية معدلة للانتشار الخلفي

وبالمثل، ولكن في الاتجاه الخلفي، تُستعمل الأخطاء المحسوبة من مخارج آخر طبقة مخفية لتعديل الأوزان في الطبقة المخفية السابقة لها، وهكذا، حتـــى تُعدَّل أوزان أول طبقة مخفية من طرف الدخل. في هذه الطريقة، تنتشر أخطاء الخرج النهائي في الاتجاه الخلفي طبقة بعد أخرى مع القيام بتعديل أو تصحيح لأوزان الطبقة الموافقة بطريقة تكرارية. تتكرر العملية مرات عديدة لكل نموذج في مجموعة التدريب حتـــى يتقارب الخطأ الكلي إلى قيمة صغرى،

أو حتـــى يتم الوصول إلى عدد محدد من مرات التكرار.

لتبسيط اشتقاق طريقة تعليم الانتشار الخلفي، سنبدأ بشبكة MLFF التي لها طبقة مخفية واحدة، ومن ثم سنقوم بتعميم النتائج لعدد من هذه الطبقات. وهكذا ستكون طبقة الدخل متصلة اتصالاً تاماً مع جميع وحدات الطبقة المخفية، وجميع هذه الوحدات متصلة مع طبقة الخرج.

وسنعتمد التدوين التالي لوسطاء الشبكة: يرمز لأوزان الوصلات بيـــن وحدة طبقة الدخل i ووحدة الطبقة المخفية j بـــ v_{ji} حيث $i=1,2,...,n$ و $j=1,2,...,h$، ويرمز لأوزان الوصلات بين وحدة الطبقة المخفية j ووحدة الخرج k بـــ w_{kj} حيث $k=1,2,...,m$ كما هو موضح في الشكل (3.6).



الشكل 3.6: وصلات الشبكة المتعددة الطبقات مع المتغيرات

يرمز لنموذج دخل التدريب p ذي البعد n بـــ $\mathbf{x}^p$ حيث $p = 1, 2, ..., P$، ويرمز لخرج الوحدة j في الطبقة المخفية لنموذج الدخل $\mathbf{x}^p$ بـــ y_j^p حيث $j=1,2,...,h$. وبالمثل يرمز لخرج الوحدة k في طبقة الخرج لنموذج $\mathbf{x}^p$ بـــ z_k^p، ويرمز للخرج المنشود بـــ t_k^p حيث $k=1,2,..., m$.

وسنستعمل أيضاً تابعَ تفعيل غيرَ خطي واحداً لكل وحدات الطبقة المخفية والخرج. وعندما لا يكون هناك التباس أو خشية اختلاط في المعطيات سنهمل رمز النموذج p لتبسيط العلاقات. سنعرّف المداخل التركيبية التالية:

$$H_j = \sum_{i=1}^{n} v_{ij} x_i$$

$$I_k = \sum_{j=1}^{h} w_{kj} y_j \qquad (2.6)$$

حيث H_j هو الدخل التركيبـي net للوحدة j من الطبقة المخفية و I_k هو الدخل التركيبي net للوحدة k من طبقة الخرج. تعطى المخارج المحسوبة بالوحدة j من الطبقة المخفية والوحدة k من طبقة الخرج كما يلي:

$$y_j = f(H_j) \quad ,i = 1,2,...,h$$

$$z_k = f(I_k) \quad ,k = 1,2,...,m \qquad (3.6)$$

حيث f تابع التفعيل التفاضلي ويُحدَّد كيفياً. وهكذا تعطى استجابة وحدة الخرج k لنموذج الدخل **x** كما يلي:

$$z_k = f(I_k) = f(\sum_{j=1}^{h} w_{kj} y_j) = f(\sum_{j=1}^{h} w_{kj} f(H_j))$$

$$= f(\sum_{j=1}^{h} w_{kj}\, f(\sum_{i=1}^{n} v_{ji}\, x_i)) \qquad (4.6)$$

وهذه المخارج موضحة في الشكل (3.6). من الواضح أنه إذا كان f تابعاً غير خطي، فإن z_k سيكون تابعاً غير خطي لنموذج الدخل **x** ووسطاء النظام **W** = (**v**, **w**)، أي **z** = **g** (f, **x**, **w**) حيث **g** تابع شعاعي ببعد m.

نعرف متوسط خطأ النظام E_{tot} كمتوسط أخطاء الخرج عبر كل أخطاء نماذج التدريب E^p كما يلي:

$$E_{tot} = \frac{1}{P} \sum_{p=1}^{P} E^p \qquad (5.6)$$

يعتمد عدد نماذج التدريب في مجموعة التدريب على المسألة المدروسة (التطبيق العملي) وعلى مصدر مجموعة التدريب. من الواضح أن P يمكن أن تكون محددة أو غير محددة؛ إذ تكون محددة في حالة مجموعة تدريب بحجم محدد P، التـــي نتوقع أن تكون عينة تمثيلية من توزيع عام ρ(**x**). بتعبير أدق سنعرف متوسط مربع الخطأ كنهاية لهذا المجموع تأخذ الشكل التالي:

$$E_{tot} = \lim_{P\to\infty} \frac{1}{P} \sum_{p=1}^{P} E^p \quad (6.6)$$

من الواضح أن خطأ النظام سينخفض إذا انخفض خطأ كل نموذج تدريب E^p لأي حجم P.

وهكذا، كما في قاعدة دلتا، سنطوّر خوارزمية لتصحيح الأوزان بحيث يجري تقليل الأخطاء بالنسبة إلى التعديل في الأوزان. وسنغير الأوزان لكل تمثيل نموذج على نحو متتابع بحيث تخفض قيم أخطاء النظام، تكرارياً، عن قيمها السابقة.

يمكن أن ينجز هذا إذا عُدِّلت الأوزان تعديلاً متناسباً مع سالب تدرج الخطأ. أي عند الخطوة s +1 من عملية التدريب، سيكون تعديل الوزن متناسباً مع مشتق الخطأ المقيس E^p المحسوب لـــ s تكراراً، ويمكن أن يعبر عنه رياضياً كما يلي:

$$\Delta \mathbf{W}(s+1) = -\eta \frac{\partial E^p}{\partial \mathbf{W}(s)} \quad (7.6)$$

حيث η عامل التعليم الثابت، و

$$\frac{\partial E^p}{\partial \mathbf{W}} = \left[\partial E^p/\partial v_{11}, \partial E^p/\partial v_{12}, ..., \partial E^p/\partial v_{hn}, \partial E^p/\partial w_{11}, \partial E^p/\partial w_{12}, ..., \partial E^p/\partial w_{hm}\right]$$

يعطى تدرج خطأ النظام بالمعادلة:

$$\frac{\partial E_{tot}}{\partial \mathbf{W}} = \frac{1}{P} \sum_{p=1}^{P} \frac{\partial E^p}{\partial \mathbf{W}} \quad (8.6)$$

يمكن أن يعرف تابع الخطأ E^p بطرق مختلفة: مثلاً كمتوسط مربع الخطأ، أو القيمة المطلقة للخطأ، ..الخ. لتطوير نموذج أولي لخوارزمية Bp، سنستعمل متوسط مربع الخطأ MSE (Mean Square Error) لأنه أحد أكثر القياسات استعمالاً (سنناقش حسنات وسيئات هذه الطريقة لاحقاً).

يعرّف مربع الخطأ لنموذج دخل p كما يلي:

$$E^p = \frac{1}{2} \sum_{k=1}^{m} (t_k^p - z_k^p)^2 \quad (9.6)$$

في عملية إيجاد التعبير المناسب لتعديل الأوزان (المعادلة (7.6)) يجب أخذ المشتق الجزئي

لـ E^p للأوزان v_{ji} و w_{kj}، وهذا يتطلب استعمال قاعدة السلسلة، لأننا نريد أن نعبر عن عملية التعديل بوسطاء النظام المتعددة: المداخل، والمخارج المحسوبة، والأوزان. عندئذ يمكن أن يجري حساب تعديلات الأوزان كما يلي:

$$w_{kj}(s+1) = w_{kj}(s) + \Delta w_{kj}$$
$$\Delta w_{kj}(s+1) = -\eta \frac{\partial E^p}{\partial w_{kj}(s)} \qquad (10.6)$$

وذلك بأخذ مشتق التابع E^p المعطى بالنسبة إلى w_{kj}. سنأخذ مشتقات الحدود التي تعتمد وظيفياً على الأوزان؛ I_k وH_j (لاحظ أننا أهملنا الدليل p في العلاقات التالية بغية التبسيط؛ أي $E = E^p$). وهكذا لتقيم الحدود على نحو منفرد مثل:

$$\partial E / \partial w_{kj}$$

في التعبير (10.6) نستعمل التعابير التالية:

$$E = \frac{1}{2}\sum_{k=1}^{m}(t_k^p - z_k^p)^2, \quad I_k = \sum_{j=}^{h} w_{kj} y_j, \qquad z_k = f(I_k) \qquad (11.6)$$

سنركز أولاً على تحديث أوزان وحدات الخرج، لذا سنستعمل الأخطاء الفعلية لإيجاد قاعدة التحديث. لدينا:

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial I_k}\frac{\partial I_k}{\partial w_{kj}} = \frac{\partial E}{\partial I_k}\left(\frac{\partial\left(\sum_{j=1}^{h} y_j w_{kj}\right)}{\partial w_{kj}}\right) \qquad (12.6)$$

الحد بين الأقواس في المعادلة (12.6) يمكن أن يعطي مباشرة قيمة:

$$\frac{\partial\left(\sum_{j=1}^{h} y_j w_{kj}\right)}{\partial w_{kj}} = y_k$$

وباستعمال قاعدة السلسلة مرة ثانية نجد:

$$\frac{\partial E}{\partial I_k} = \frac{\partial E}{\partial z_k}\frac{\partial z_k}{\partial I_k} = -(t_k - z_k).f'(I_k) \qquad (13.6)$$

من (11.6) نجد أنه في حالة وحدة الخرج k يكون لدينا:

$$\partial E/\partial z_k = -(t_k - z_k) \quad , \quad \partial z_k/\partial I_k = f'(I_k) \qquad (14.6)$$

إذا عرّفنا الآن:

$$\delta_k = (t_k - z_k)\cdot f'(I_k) \qquad (15.6)$$

فنستطيع كتابة قاعدة تحديث أوزان وحدات الخرج كقاعدة دلتا لـ Widrow-Hoff على النحو التالي:

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} = \eta\,\delta_k\,y_j \qquad (16.6)$$

تطبّق هذه القاعدة على كل أوزان وحدات الطبقة المخفية المتصلة مع طبقة الخرج في حالة تمثيل نموذج مفرد.

إذا عُدنا إلى أوزان الطبقة المخفية υ_{ji} فسنرى أنه في هذه الحالة ليس لدينا قيم منشودة تحسب منها الأخطاء. عوضاً عن ذلك، علينا ــ بطريقة ما ــ استعمال الأخطاء من وحدات الخرج لتعديل الأوزان بين الدخل والطبقة المخفية.

لذا سنستعمل قاعدة السلسلة تكرارياً لربط أخطاء الخرج مع هذه الأوزان. لاحظ أن هذه الأوزان (المعادلة (4.6)) متوغلة بعمق في تابع الخطأ. نحتاج الآن إلى التعبير عن:

$$\Delta \upsilon_{ji} = -\eta \frac{\partial E}{\partial \upsilon_{ji}} = -\eta \frac{\partial E}{\partial H_j}\cdot\frac{\partial H_j}{\partial \upsilon_{ji}} \qquad (17.6)$$

لاحظ أن المشتق الجزئي الأخير يمكن أن يقيم مباشرة من المعادلة (2.6) أي :

$$\frac{\partial H_j}{\partial \upsilon_{ji}} = \sum_i \frac{\partial}{\partial \upsilon_{ji}}(\upsilon_{ji}x_i) = x_i$$

للحصول على الحل للمشتق الأول في (17.6) سنستعمل قاعدة السلسلة للحصول على:

$$\partial E/\partial H_j = \partial E/\partial y_j \cdot \partial y_j/\partial H_j = \partial E/\partial y_j\; f'(H_j) \qquad (18.6)$$

باشتقاق $\partial E/\partial y_j$ مباشرة (من المعادلات (4.6) و(9.6)) نحصل على:

$$\frac{\partial E}{\partial y_j}=\frac{1}{2}\sum_k \frac{\partial (t_k - f(\sum_j w_{kj} y_j))^2}{\partial y_j}$$
$$=-\sum_k (t_k - z_k)\cdot f'(I_k) w_{kj}$$

نستطيع الآن استنتاج قاعدة التحديث لوحدات الطبقة المخفية كما يلي:

$$\Delta \upsilon_{ji} = -\eta \frac{\partial E}{\partial \upsilon_{ji}} = -\eta \frac{\partial E}{\partial H_j} \frac{\partial H_j}{\partial \upsilon_{ji}}$$
$$= -\eta \frac{\partial E}{\partial H_j} x_i = -\eta \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial H_j} x_i$$
$$= -\eta \frac{\partial E}{\partial y_j} x_i f'(H_j)$$

$$= -\eta\, x_i \left[-\sum_k (t_k - z_k) f'(I_k) w_{kj} \right] f'(H_j), \quad \delta_k = (t_k - z_k) f'(I_k)$$
$$= \eta\, x_i f'(H_j) \sum_k \delta_k w_{kj}, \qquad \delta_j = f'(H_j) \sum_k \delta_k w_{kj}$$

$$\Delta \upsilon_{ji} = \eta\, \delta_j x_i \qquad (19.6)$$

لإجمال ما سبق سنكرر قاعدتـــي التحديث لوحدات طبقة الخرج والطبقة المخفية:

وحدات الخرج:

$$\Delta w_{kj} = \eta \delta_k y_j = \eta (t_k - z_k) f'(I_k) y_j$$
$$\delta_k = (t_k - z_k) f'(I_k) \qquad (20.6)$$

الوحدات المخفية :

$$\Delta \upsilon_{ji} = \eta \delta_j x_i = \eta\; x_i f'(H_j) \sum_k \delta_k w_{kj}$$
$$\delta_j = f'(H_j) \sum_k \delta_k w_{kj} \qquad (21.6)$$
$$\delta_k = (t_k - z_k) f'(I_k)$$

في البداية، وقبل تطبيق عملية التدريب بالانتشار الخلفي، تعطى مصفوفات الأوزان قيماً

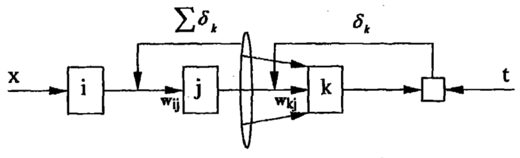
أوليةً: أعداداً حقيقية أو كيفية صغيرة جداً. ومن ثم تقدم أشعة نموذج التدريب $\mathbf{x}^p$ حيث $p = 1, 2, .., P$ إلى الشبكة. بعد زمن معين يكون قد تم حساب نموذج خرج الشبكة الموافق $\mathbf{z}^p$ لنموذج الدخل. يقارن خرج الشبكة المحسوب مع نموذج الخرج المنشود $\mathbf{t}^p$ فتُعرَف أخطاء النموذج وهي $(t_k^p - z_k^p)$ حيث $k = 1,2,...,m$. تُرسل هذه الأخطاء بعدئذ بالاتجاه الخلفـــي لحساب تعديل الأوزان كما هو مبين سابقاً (المعادلات (20.6) و(21.6)).

لفهم عملية التعليم على نحو أفضل علينا فحص معادلات الانتشار الخلفي، وسنبدأ من حسابات الطبقة الأخيرة.

في مسار الحساب بالاتجاه الأمامي يجري تكوين الجداءات من خرج كل عقدة في الطبقة المخفية y_j مع الأوزان w_{kj} الموافقة المتصلة مع الوحدة k لطبقة الخرج، ومن ثم تجمِّع هذه الوحدة الجداءات المثقلة لتكوين دخلها net وإعطاء تفعيل خرجها z_k. بعدئذ يُحسب الخطأ $(t_k^p - z_k^p)$ ليُستعمل في العلاقة (15.6) في عملية الانتشار الخلفي لتعديل الأوزان w_{kj} المتصلة مع الوحدة k. تنجز هذه العملية في كل وحدة طبقة خرج $k = 1,2,...,m$ وفق المعادلة:

$$w_{kj}^{new} = w_{kj}^{old} + \Delta w_{kj} = w_{kj}^{old} + \eta y_j (t_k - z_k) f'(I_k) \qquad (22.6)$$

ولحساب الأوزان v_{ji} من الدخل إلى الطبقة المخفية، لا تتوفر قيم هدف لاستعمالها في حساب الأخطاء مباشرة. وعوضاً عن ذلك، تستعمل أخطاء الخرج وتوزعها بطريقة ما لتعديل الأوزان الواصلة عقد الدخل i إلى عقد الطبقة المخفية j.



الشكل 4.6: عملية تعديل وزن الانتشار الخلفي

من أجل ذلك نستعمل قيم δ_j كأخطاء لكل وحدة مخفية j كما يلي:

$$\upsilon_{ji}^{new} = \upsilon_{ji}^{old} + \Delta\upsilon_{ji} = \upsilon_{ji}^{old} + \eta x_i f'(H_j)\sum_k \delta_k w_{kj} \qquad (23.6)$$

إن حدود δ_k و δ_j في العلاقة (23.6) مختلفة بعضها عن بعض وقد عرفت في المعادلات (20.6) و (21.6). يوضح الشكل (4.6) السابق عملية التعديل لطبقتين.

4.6 التدريب المؤجل بأسلوب الدفعة الواحدة

Off-line or Batch process training

عوضاً عن تعديل الأوزان بعد معالجة كل نموذج من نماذج الدخل يمكن تكديس الأخطاء E^p عبر كل مجموعة التدريب $\{\mathbf{x}^p, \mathbf{t}^p | p = 1,2,...,P\}$ ، ومن ثم القيام بالتعديلات. يعرف هذا الإجراء بتحديث الدفعة الواحدة أو التحديث المؤجل حيث ينفذ التحديث للدور بأكمله؛ أي مرور كامل خلال مجموعة التدريب.

تخزن الأخطاء لكل نموذج في مجموعة التدريب، وبعد دور التدريب يُحسب الخطأ الكلي $E_{tot} = \sum_p E^p$ وتعدَّل الأوزان وفقاً للأخطاء المكدسة. وهذا يعنـــي أن $\sum_p w_{kj}$ و $\sum_p \upsilon_{kj}$ تكون ضرورية لخوارزمية تحديث الانتشار الخلفي(المعادلة (8.6)). تتكرر هذه العملية حتـــى إتمام التدريب.

تعتبر طريقة تعديل الأوزان بعد تقديم كل نموذج عوضاً عن تعديلها بعد كل دور من أكثر الطرائق استعمالاً. على أي حال، يمكن أن يكون تدريب الدفعة لبعض التطبيقات أكثر فعالية.

يلزم لتنفيذ التدريب المؤجل أن تكون مجموعة التدريب كاملة قبل بداية عملية التدريب. وهذا الشرط قد يكون محققاً في بعض التطبيقات مثل تطبيقات الزمن الحقيقي (عملية التحكم) أو توازن الإشارة المتكيف (Widrow & Stearnes عام 1985[54]).

يمكن أن تعمم قاعدة تحديث الانتشار الخلفي على أي عدد من الطبقات، وسيكون لها الشكل العام التالي :

$$\Delta w_{ji} = \eta \sum_p \delta_{out} H_{in} \qquad (24.6)$$

حيث تشير الأدلة in وout إلى إشارات دخل وخرج الوحدة المرافقة لوحدة معطاة، وترمز i وj إلى الوصلات من الوحدة i إلى الوحدة j. سيتغير شكل δ وذلك على حسب الطبقة التـــي يطبق فيها. في حالة وحدة الطبقة الخارجية سيكون لها الشكل المعطى بالمعادلة (15.6)، أما وحدات الطبقات المخفية فسيكون لهذا التابع الشكل المعطى بواسطة المعادلة (19.6). لتأكيد من هذه النتائج، نطبّق قاعدة السلسلة تكرارياً في حالة الطبقات المخفية المتعددة. مثلاً، إذا كان للشبكة Q طبقة، فإن قاعدة التحديث لأوزان الطبقة رقم q ستكون:

$$\Delta w_{ji}^{q} = \eta \delta_i^q O_j^{q-1} \quad (25.6)$$

حيث O_j^{q-1} هو خرج الوحدة رقم j من الطبقة q-1، و δ^q لها الشكل العام:

$$\delta_i^q = f'(H_i^{q-1}) \sum_j \delta_j^{q+1} w_{ji}^q \quad (26.6)$$

وهذا يعني: حاصل جداء المشتق لخرج الوحدة التـــي عدلت أوزانها بالمجموع المثقل لكميات دلتا الخرج المرافقة للوحدة. تحسب كميات دلتا بسهولة عندما تنتشر الأخطاء في الاتجاه الخلفي تتابعياً طبقة بعد طبقة.

بسبب أهمية خوارزمية تعليم الانتشار الخلفي سنقوم بتلخيص هذه الخوارزمية فيما يلي من أجل شبكة عصبونية متعددة الطبقات بتغذية أمامية بعدد كيفي من الطبقات Q حيث q = 1, 2, .., Q. وسنرمز بـــ H_i^q و Q_i^q لمداخل ومخارج الوحدة رقم I من الطبقة رقم q، و w_{ji}^q هو وزن الوصلة بين الوحدة رقم I في الطبقة q-1 والوحدة رقم j في الطبقة q.

خوارزمية الانتشار الخلفي Backpropagation Algorithm:

1. إعطاء الأوزان قيماً أولية، أي قيماً كيفية صغيرة ضمن المجال [λ- ,λ]، حيث قيمة λ > 0 صغيرة .

2. اختيار، زوج من نماذج التدريب $\{x^p, t^p\}$، عشوائياً وحساب قيم الخرج بالاتجاه الأمامي لكل وحدة j من كل طبقة q، وهكذا:

$$O_j^q = f(\sum_i O_i^{q-1} w_{ji}^q)$$

لاحظ أن مداخل الطبقة الأولى ذات دليل صفري أي: $O_j^0 = x_j$ والأوزان $w_{ji}^0 = 1$.

3. استعمال القيم O_j^Q المحسوبة بواسطة وحدات الطبقة النهائية وقيم الهدف الموافقة t_j^p لحساب كميات دلتا:

$$\delta_j^Q = (O_j^Q - t_j^p) f'(H_j^Q)$$

لكل j باستعمال النموذج p.

4. حساب كميات دلتا لكل الطبقات السابقة بواسطة الانتشار الخلفي باستعمال الأخطاء:

$$\delta_j^{q-1} = f'(H_j^{q-1}) \sum_i \delta_i^q w_{ji}^q$$

لكل j في كل الطبقات $q = Q, Q-1, .., 2$

5. تحديث كل الأوزان باستعمال:

$$w_{ji}^{new} = w_{ji} old + \Delta w_{ji}^q$$
$$\Delta w_{ji}^q = \eta \delta_i^q O_j^{q-1}$$

لكل طبقة q.

6. عودة إلى الخطوة 2 والتكرار لكل نموذج p حتـــى يصل الخطأ الكلي إلى مستوى مقبول.

ومع أن الخطوات الرئيسية للخوارزمية أصبحت واضحة، فإن أسئلة عديدة ستطرح ويجب الإجابة عنها قبل تعيين الشبكة التـــي ستستخدم لتطبيق معطى:

1. ما هي الوظائف (التطبيقات) التـــي تستطيع الشبكات العصبونية الصنعية المتعددة الطبقات بتغذية أمامية تحقيقها ؟
2. ما هو عدد الطبقات المستعملة؟ وما هو عدد الوحدات الأمثلي في كل طبقة من التطبيق المطلوب لأداء المهمة المعطاة؟
3. ما هي خطة ترميز الدخل المستخدمة لنماذج الدخل؟
4. هل ستكون هذه النماذج معيارية بطريقة ما؟
5. ما هي سعة تخزين الشبكة المعطاة؟
6. هل بنية التغذية الأمامية المتصلة كلياً هي أفضل البنى للتطبيق؟ أم ستعتبر مخططات توصيل اختيارية، مثل الشبكات الموصلة جزئياً بطبقات ثلاثية البعد أو شبكات موصلة على نحو

مبعثر؟

7. هل ستستخدم طبقات هجينية ، أي طبقات مختلفة من شبكات MLFF مع Bp وشبكة أخرى(Kohonen, ART, Hamming , RCE)؟
8. ما هي أفضل خوارزمية لتعليم الشبكة؟
9. ما هي تركيبة مجموعة التدريب الجيدة؟
10. ماهو حجم مجموعة نماذج التقييم والاختبار لمسألة معطاة؟
11. كيف يمكن أن تتطور مجموعة التدريب؟
12. هل سيكون اختيار نماذج التدريب نظامياً أم عشوائياً خلال التدريب؟
13. هل سيضاف الضجيج إلى نماذج التدريب؟
14. هل سينجز تعديل الأوزان بعد كل نموذج أم بعد كل دور تدريب؟
15. كيف سيُكتَشف ويُتجنَّب الأصغر المحلي لتابع الخطأ؟
16. كيف يمكن تسريع عملية التعليم؟
17. كيف ستعطى القيم الأولية للأوزان؟
18. ما هي توابع التفعيل الفُضْلى للمسألة المعطاة؟
19. ما هو التعميم؟
20. كيف يمكن للشبكة أن تتعلم التعميم؟
21. ما هي مميزات الإنجاز المرغوب بها (الدقة، الموثوقية، المقدرة على التعميم) للمسألة المعطاة؟
22. كيف تختبر الشبكة ويقاس أداؤها؟
23. كيف تتحسن سلوكيات الشبكات الصغيرة لزيادة حجمها لتصبح شبكات ضخمة؟

5.6 توابع تفعيل الانتشار الخلفي

لقد ذكرنا من قبل أن خوارزمية الانتشار الخلفي تتطلب توابع تفعيل تفاضلية محددة. يعتبر التابع sigmoid عموماً (والمعروف بالتابع النسبي logistic) المحقق لهذه المتطلبات أكثر توابع التفعيل المستعملة في خوارزمية الانتشار الخلفي.

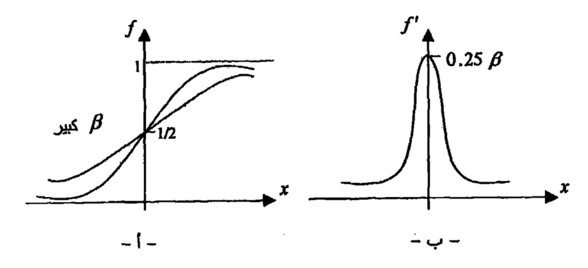
عرّف هذا التابع على **R** وتكون قيمته محددة بين الصفر والواحد، وهو تابع متزايد ومستمر بشكل الحرف **S**، وله الصيغة العامة التالية:

$$f(x) = \frac{1}{1+\exp(-\beta x)} \quad (27.6)$$

حيث يحدد β انحدارية المنحنـــي ذي الشكل **S** المبين في الشكل (5.6 أ) بقيم مختلفة لـــ β . يعطى مشتق هذا التابع الموضح في الشكل(5.6 ب) بالعلاقة التالية:

$$f'(x) = \beta f(x)(1-f(x)) \quad (28.6)$$

لاحظ أن المشتق له قيمة عظمى تساوي 0.25β عند النقطة $x=0$. وكلما ابتعدت قيمة x عن الصفر فإن المشتق يقترب من الصفر.



الشكل 5.6: مخطط ـــ أ ـــ تابع Sigmoid وـــ ب ـــ مشتقه.

تعطـــى قواعد تحديث الانتشار الخلفي عند استعمال تابع التفعيل sigmoid من العلاقات (20.6) و(21.6) كما يلي:

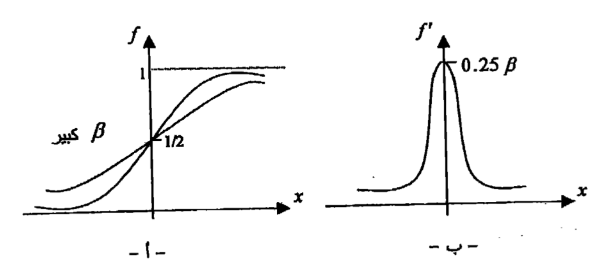
$$\begin{aligned} \delta_k &= O_k(1-O_k)(t_k-O_k) \\ \delta_j &= O_j(1-O_j)\sum_k w_{kj}\delta_k \end{aligned} \quad (29.6)$$

هناك تابع آخر يستعمل كثيراً في هذه الخوارزمية هو تابع الظل القطعي. هذا التابع له شكل مشابه لتابع sigmoid؛ فهو متزايد ومستمر ولكن مجال قيمه من -1 حتـــى +1 عوضاً عن [0, 1]. يعطى هذا التابع الموضح في الشكل (6.6 أ) بالعلاقة التالية:

$$f(x)=\frac{(e^{x}-e^{-x})}{(e^{x}+e^{-x})} \qquad (30.6)$$

وبإجراء تغير للمتحول وتعويض βx عوضاً عن x نحصل على وسيط نستطيع بواسطته تعديل انحدارية تابع الظل القطعي كما في تابع sigmoid. إن مشتق هذا التابع له الشكل البسيط التالي، الموضح في الشكل (6.6 ب):

$$f'(x)=\beta(1-f(x))^{2} \qquad (31.6)$$



الشكل 6.6: مخطط ـــ أ ـــ تابع الظل القطعي وـــ ب ـــ مشتقه.

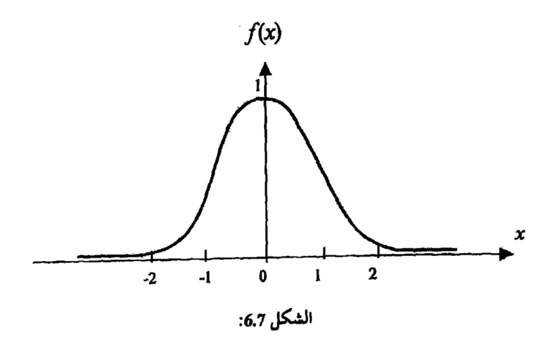
تعطى قواعد تحديث الانتشار الخلفي لهذا التابع من(20.6) و(21.6)كما يلي:

$$\begin{aligned}\delta_k&=(1-O_k^2)(t_k-O_k)\\ \delta_j&=(1-O_j^2)\sum_k w_{kj}\delta_k\end{aligned} \qquad (32.6)$$

يمكن أن تستعمل توابع تفعيل أخرى على نحو فعال في شبكات التغذية الأمامية متعددة الطبقات مثل التوابع اللوغارتمية (تابع الأنتروبـــي المعطى في الفصل الثالث)، والتوابع المثلثية (الجيب والتجيب)، وتوابع الأساس الشعاعي (لغوص).... الخ.

يعطى تابع غوص الموضح في الشكل (7.6) بالعلاقة التالية:

$$f(x)=\exp(-x^{2}) \qquad (33.6)$$



الشكل 6.7:

ويعطى مشتقه بالعلاقة التالية:

$$f'(x) = -2xf(x) \tag{34.6}$$

استعمل Hecht-Nielson عام 1990[46] تابع التماثل الخطي $f(net) = net$ كتابع تفعيل على وحدات الخرج، وخاصة إذا كانت قيم الهدف مستمرة بدلاً من أن تكون ثنائية أو ثنائية القطبية.

في بعض التطبيقات لا يكون الإشباع مفيداً بل قد يكون ضاراً أحياناً، لذا يمكن استعمال تابع تفعيل غير قابل للإشباع. من الأمثلة المناسبة لهذه التوابع:

$$f(x) = \begin{cases} \log(1+x) & x > 0 \\ -\log(1-x) & x < 0 \end{cases} \tag{35.6}$$

لاحظ أن مشتق هذا التابع مستمر عند $x = 0$:

$$f'(x) = \begin{cases} \dfrac{1}{1+x} & x > 0 \\ \dfrac{1}{1-x} & x < 0 \end{cases} \tag{36.6}$$

في بعض التطبيقات يمكن تركيب هذا التابع مع تابع التماثل الخطي على وحدات الخرج.

مثال 1:

شبكة أمامية التغذية متعددة الطبقات من أجل ضغط المعطيات،

يمكن أن تستعمل شبكات MLFF لضغط المعطيات وذلك بتدريب الشبكة لتتصرف كشبكة ترافق ذاتي (شعاع دخل التدريب يساوي شعاع الخرج المنشود) بوحدات مخفية أقل من عدد وحدات الدخل والخرج (Cottrell, Munro, Zipser عام 1989[55]).

في هذا المثال البسيط اختيرت مجموعة معطيات الصورة كمجموعة الأحرف A, B, C, D, E, F, G, H, I, J الموضحة في الشكل (8.6)، حيث عرِّف كل حرف بقيم ثنائية على شبكية مخططة بحجم 7× 9 عنصر صورة (Pixels).

ولما كان كل حرف مُثِّل في معطيات الدخل بشعاع له 63 مركبة ثنائية، فإن طبقة الدخل للشبكة العصبونية سيكون لها 63 وحدة دخل. وقد أُعطيت الطبقة المخفية عدداً أصغر من الوحدات بغية عملية الضغط.

من المعروف أن مجموعة من N نموذجَ دخلٍ متعامداً يمكن أن تطبق على $\log_2 N$ وحدة مخفية لتكوين رمز ثنائي بنموذج مختلف لكل N نموذج دخل (مجموعة البحث العلمي Rumelhart, McClelland, PDP عام 1986 [56]).

ولما كانت الأحرف في مجموعة نماذج التدريب ليست متعامدة، فإن قيمة $\log_2 N$ يمكن أن تؤخذ كحد أدنى نظرياً فيما يتعلق بعدد الوحدات المخفية التـــي يمكن أن تستعمل لعملية الضغط إذا كان المطلوب هو إعادة تشكيل الحروف تامة، وهذا ما يسمى بإعادة تشكيل عدم الخسارة.

إن عدد الوحدات المخفية سيكون متغيراً كمهمة جزئية من المسألة المدروسة وعدد وحدات الدخل سيكون 63 وحدة (لإعادة التشكيل). ونقول عن الشبكة إنها تعلمت نموذجاً إذا كانت قيمة الخرج المحسوبة ضمن مجال التسامح المخصص للقيم المنشودة (0 أو 1).

النموذج الأول	النموذج الثاني	النموذج الثالث	النموذج الرابع	النموذج الخامس
00 000	00	00 0	000	0
000 000	0 000 0	0 000 0	0 00 00	0 000 0
000 000	0 000 0	000000	0 000 0	0 0 000
00 0 00	0 00	000000	0 000 0	0 000
00 0 00	0 000 0	000000	0 000 0	0 0 000
0 0	0 000 0	000000	0 000 0	0 00000
0 000 0	0 000 0	0 000 0	0 00 00	0 000 0
0	00	00 00	000	0
0000000	0000000	0000000	0000000	0000000

النموذج السادس	النموذج السابع	النموذج الثامن	النموذج التاسع	النموذج العاشر
0	00 0	0	0 0	00 0
0 000 0	0 000 0	0 000 0	000 000	0000 00
0 0 000	000000	0 000 0	000 000	0000 00
0 000	000000	0 0	000 000	0000 00
0 0 000	000	0 000 0	000 000	000 00
0 00000	0000 0	0 000 0	000 000	000 00
0 00000	0 000 0	0 000 0	000 000	000 00
000	00 00	0	0 0	0 000
0000000	0000000	0000000	0000000	0000000

الشكل 8.6: نماذج التدريب

يبين الجدول التالي عدد الأدوار اللازمة لكي تتعلم الشبكة نماذج الدخل العشرة من أجل قيمتين للتسامح هما [0.1, 0.2].

	المجال				
	0.1	0.2		0.1	0.2
عدد الوحدات	عدد الأدوار	عدد الأدوار	عدد الوحدات	عدد الأدوار	عدد الأدوار
7	660	540	19	180	90
8	700	210	20	180	90
9	590	270	21	160	175
10	415	210	22	150	70
11	375	230	23	135	60
12	235	140	24	140	340
13	235	110	25	135	55
14	200	95	26	135	70

55	130	27	110	270	15
60	120	28	90	210	16
65	140	29	70	200	17
70	125	30	-	200	18

جدول (6.1) عدد الأدوار اللازمة كتابع لعدد الوحدات المخفية

بعبارة أخرى، تعتبر استجابة الوحدة غير صحيحة إذا كان لها تفعيل لا يزيد على 0.2 وعنصر الصورة الموافق (في نموذج التدريب) لتلك الوحدة كان "Off". وبالمثل، تعتبر استجابة الوحدة صحيحة، الموافقة لعنصر الصورة الذي كان "On"، إذا كان لها تفعيل لا يقل عن 0.8.

وكذلك تعتبر استجابة الوحدة غير صحيحة إذا كان لها تفعيل ليس أقل من 0.1 وعنصر الصورة الموافق (في نموذج التدريب) لتلك الوحدة كان "Off". وتعتبر استجابة الوحدة صحيحة، الموافقة لعنصر الصورة الذي كان "On"، إذا كان لها تفعيل لا يقل عن 0.9. إن كل الوحدات في الشبكة كان لها تفعيل صحيح (ضمن المجال المخصص) لكل نماذج التدريب قبل أن يعتبر التعليم ناجحاً. من الأهمية بمكان ملاحظة أن دقة هذه النتائج كانت بقيمة 100% للأحرف التـــي أعيد تشكيلها في مجموعة التدريب.

في بعض تطبيقات الاتصالات، تكون هذه الدقة ضرورية لكي تكون عملية إعادة التشكيـــل عديمة الخسارة محققة. فـــي تطبيقات أخـــرى، كبعض أنـــواع معـــالجة الصور. (Arozuggah & Namphol عام 1990 [57]، Cottrell & munro & Zipser عام 1989 [55]، Sonehara & Kawato & miyake عام 1989 [58]) يمكن التسامح بمستوى ما من الضجيج في النموذج المسترد.

النتائج المعروضة في الجدول السابق أعطيت في حالة مجموعة واحدة من الأوزان الأولية لكل بنية وكل مجال تسامح. وكما هو ملاحظ من النتائج فقد احتجنا إلى عدد كبير من الأدوار في حالتـــي 21 و24 طبقة مخفية. باستعمال مجال تسامح كبير، يُعكس التغير في مجموعة التدريب لأوزان أولية مختلفة. نلاحظ أيضاً تقارب الخوارزمية لــ 18 وحدة مخفية ومجال تسامح 0.2.

مثال 2:

شبكة MLFF بالانتشار الخلفي لإعطاء توابع الجيب،

دربت شبكة عصبونية أمامية التغذية متعددة الطبقات لتقوم بتطبيق أشعة الدخل (x_1, x_2) على قيمة خرج موافقة y كما يلي:

- تقع قيم الدخل (x_1, x_2) بين الصفر والواحد بخطوة 0.2.
- تعطى قيم الخرج المنشود الموافقة بواسطة $y = \sin(2\pi x_1)\cdot\sin(2\pi x_2)$

هذه المهمة صعبة جداً على خوارزمية الانتشار الخلفي التقليدية، لذا سنستعمل توابع تفعيل لوغارتمية للوحدات المخفية (الشبكة لها طبقة مخفية واحدة بعشر وحدات) وتابع التماثل الخطي لوحدات الخرج، مع معدل تعليم يساوي 0.05.

لقد أنجزت الشبكة متوسط مربع خطأ يساوي 0.024 في 5000 دور. والنتائج معطاة في الجدول التالي، وأعطيت قيم الخرج المنشود بالخط المائل وقيم الخرج المحسوب بالشبكة بالخط الغامق.

6.6 علامات اختبار إنجاز الشبكات المتعددة الطبقات بتغذية أمامية

سنصف في هذه الفقرة بعض العلامات والمؤشرات التـي هي أكثر استعمالاً من قبل الباحثين لاختبار مقدرات الإنجاز المتنوعة للشبكات العصبونية الصنعية المتعددة الطبقات الأمامية التغذية والتوثق من النتائج النظرية بواسطة تجارب المحاكاة.

1. **XOR أو التكافؤ (التطابق) parity**: دربت هذه الشبكات لإعطاء خرج يساوي الواحد (الصفر) عندما يكون للدخل الثنائي عدداً فردياً (زوجياً) من خانات الواحد. في مسائل التكافؤ العامة، يكون هناك دخلاً ثنائياً وخرجاً ثنائياً واحداً، (مثلاً لمسألة XOR بسيطة سيكون n = 2). سيكون عدد الطبقات المخفية وعدد الوحدات مختلفاً، ولكن عادةً، تستعمل طبقة مخفية واحدة .

x_2						
1.0	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	-0.01	**-0.05**	**-0.12**	**-0.09**	**-0.01**	**-0.00**
0.8	*0.00*	*-0.90*	*-0.56*	*-0.56*	*0.90*	*0.00*
	-0.03	**-0.91**	**-0.55**	**-0.53**	**-0.97**	**-0.02**
0.6	*0.00*	*-0.56*	*-0.36*	*0.36*	*0.56*	*0.00*
	-0.00	**-0.59**	**-0.32**	**-0.37**	**-0.51**	**-0.03**
0.4	*0.00*	*0.56*	*0.36*	*-0.36*	*-0.56*	*0.00*
	-0.01	**-0.57**	**-0.33**	**-0.35**	**-0.55**	**-0.00**
0.2	*0.00*	*0.90*	*0.56*	*-0.56*	*-0.90*	*0.00*
	-0.02	**-0.84**	**-0.57**	**-0.57**	**-0.89**	**-0.01**
0.0	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*	*0.00*
	-0.02	**-0.02**	**-0.02**	**-0.02**	**-0.03**	**-0.02**
	0.0	0.2	0.4	0.6	0.8	1.0 x_1

2. **الترميز وفك الترميز Encoder /Decoder**: تكون هذه الشبكات متصلة داخلياً اتصالاً كاملاً، وتحوي عادةً طبقة دخل وخرج تحوي n وحدة وطبقة مخفية لها m وحدة، حيث $0 < m \leq n$. دخل الشبكة هو n نموذجاً ثنائياً مختلفاً، في كل منها مكان خانة واحدة مختلفة on (واحد) وباقي المداخل n-1 جميعها off (صفر).

دربت الشبكة لإعطاء نموذج الخرج الثنائـــي المماثل لنموذج الدخل. من الملاحظ أن هذا سيجبـــر الوحدات المخفية لترميـــز النماذج. مثلاً، إذا كانت قيمة n = 16 و m = 4 فـــإن الشبكـــة يجب أن تـــوفر التـــرميز $n\log_2$ الثنائـــي المكافـــئ فـــي الطبقة المخفية $(m = \log_2 n = \log_2 2^4 = 4)$. عندما تكون m صغيرة مقارنة مع n، ستكون مهمة التعلم أصعب منها في حالة m قريبة من n. وهذا يشار له أحياناً بـــ بالترميز المتوتر.

3. **متمم الترميز**: دربت هذه الشبكات لتعلم متمم مسألة الترميز/فك الترميز السابقة. وهكذا سيكون n نموذج دخل وخرج (متطابقة) جميعها on (واحد) ماعدا وحدة واحدة ستكون off (صفر). وكما في حالة مسألة الترميز وفك الترميز، يجب أن تتعلم وحدات الطبقة المخفية ترميز نماذج الدخل.

4. **انتخاب الأكثرية Majority**: دربت شبكات انتخاب الأكثرية على إعطاء خرج قيمته

الواحد عندما تكون قيمة أكثر من نصف المداخل الثنائية تساوي الواحد (on)، وإلا فإن مخارج الشبكة ستكون جميعها بقيمة الصفر. عادة تحوي الشبكات عدداً فردياً n من المداخل وخرجاً وحيداً. سيكون عدد الوحدات المخفية مختلفاً، لكنه عادةً أقل من عدد المداخل.

5. **التوافق العشوائي:** دربت هذه الشبكات باستعمال أزواج من النماذج المختارة عشوائياً، واحد للدخل ونموذج مستقل للخرج. هذه المسألة واحدة من أصعب مسائل التطبيق (mapping) للتعلم، لأنه لن يكون هناك عموماً قطاعات من النماذج المشتركة بمعالم وصفات متشابهة.

6. **تعرف الأحرف:** يُستعمل نوع شائع من علامات ودلائل اختبار مقدرة الشبكات في مسائل تعرف الأحرف. سيختلف عدد المداخل والمخارج وفقاً لمجموعة الأحرف المستعملة في مسألة التعرف وانتقاء معالم الدخل المختارة.

في أبسط الحالات، تختار بعض الأحرف المطبوعة مثل الأرقام العربية الصحيحة من الصفر حتـــى التسعة، من تشكيلة أو تشكيلات مختلفة. أما في الحالات التـــي هي أصعب، ستكون هناك محاولات لتعرف الأحرف المكتوبة يدوياً بسرعة.

لمثل هذه الاختبارات يمكـــن أن تكـــون المداخل عناصر صورة (pixels) منفـــردة (مثل 16×16) أو عدداً محدداً من المعالم المختارة بعناية، ووحدات الخرج ستوافق عادة الأحرف على نحو منفرد. سيكون عدد الطبقات المخفية مختلفاً وفقاً لنوع المسألة المنفذة، لكن عادةً، لا نحتاج إلى أكثر من طبقتين مخفيتين .

7. **التكـــتل Chunking:** تتعـــلم الشبكة في هذه المسألة تعرّف عدد الكتل أو التجمعات لـــلخانات المتشابهة (أصفار أو واحدات)؛ أي مجموعات الخانات المستمرة في سلسلة من الأرقـــام الثـــنائية. يمكـــن أن تعتبر السلسلة حلقة دائرية؛ أي الخانة في أقصى اليمين من السلسلة تعامل وكأنها مجاورة للخانة في أقصى اليسار.

مثلاً، يمكن أن تُدرَّب الشبكة لتعرُّف سلاسل الخانات الدائرية التـــي تحتوي على أكثر من كتلتين، ففي حالة السلسلة التالية 1100000111011 سيكون خرج الشبكة سالباً (إذ

هناك فقط كتلتين) وفي حالة السلسلة الدائرية 001111001111011 سيكون الخرج موجباً. وستكون الشبكة قادرةً على تعلم تعرُّف كتل من أرقام أخرى كذلك.

8. **عدد خانات الدخل:** في هذه المسألة، يجب أن تتعلم الشبكة حساب عدد خانات الواحد (الصفر) في سلسلة الدخل الثنائي. وسيكون للشبكة فقط خرج موجب واحد عندما لا يكون في الرقم المقدم للشبكة أي خانة صفر أو واحد، أما إذا قدم خانة واحدة قيمتها تساوي الواحد فسيكون هناك خرجان موجبان للشبكة وهكذا. وفي حالة n مدخلاً، يجب أن يكون للشبكة n +1 مخرجاً.

7.6 أسطح الخطأ وخواص التقارب

Error surfaces and convergence properties

عرضنا سابقاً بنية التدريب بالانتشار الخلفي على حسابات تدرج الهبوط . تبحث هذه العملية تكرارياً عن مجموعة الأوزان *W التـــي تحقق تابع خطأ أصغري عبر كل أزواج نماذج التدريب، أي:

$$E(\mathbf{W}^*) = \min_w \{\sum_{p=1}^{P} E^p(\mathbf{W}, \mathbf{x}^p)\} \qquad (37.6)$$

وهذا نوع تقليدي من مسائل الاستثمال. في بعض المسائل، يعرف التابع الموضوعي (الكلفة) عادة ليكون أعظمياً أو أصغرياً بالنسبة إلى مجموعة من الوسطاء. في هذه الحالة، ستكون قيم الأوزان الليفية *W هي الوسطاء التـــي ستحقق استمثال (أصغرية) تابع الخطأ E عبر مجموعة أزواج النماذج $\{(\mathbf{x}^p, \mathbf{t}^p) | p = 1,2,...,P\}$. في خوارزمية الانتشار الخلفي، تكون الأصغرية لتابع كلفة هي متوسط مربع الخطأ MSE (Mean-Square-Error).

إن استعمال تابع متوسط مربع الخطأ ليس طريقة أساسية (ولا مفضلة)في الانتشار الخلفي، وقد عَرَّفنا هنا تابع مربع الخطأ ليكون واحداً من التوابع المقبولة الأخرى التـــي تقيس تقريب خطأ الشبكة، ومن ثَم تنقص قيمتها كلما انخفض الفرق بين المخارج المحسوبة والمنشودة. مثلاً، يمكن أن نستعمل الخطأ المطلق، أو الخطأ الأعظمي، أو متوسط الخطأ، أو الخطأ الوسطي، أو الأنتروبـــي وتوابع أخرى عديدة.

سيعتمد اختيار قياس الخطأ على الطريقة التـــي نرغب فيها بتحديد الإنجاز، وكيفية

تدريب الشبكة. ففي طريقة التعليم بالانتشار الخلفي، يلزم أن يكون قياس الخطأ المختار تفاضلياً وينتهي إلى الصفر كلما انخفض الفرق الكلي بين النماذج المحسوبة والمنشودة (t^p-z^p) عبر مجموعة التدريب بالكامل. على أي حال، يكون قياس متوسط مربع الخطأ المعطى بواسطة:

$$E_{tot} = \frac{1}{P}\sum_{p=1}^{P} E^p \tag{38.6}$$

حيث

$$E^p = \frac{1}{2}\sum_{k=1}^{m}\left(t_k^p - z_k^p\right)^2 \tag{39.6}$$

هو تابع الخطأ الأكثر شعبية من غيره.

يستخدم هذا التابع غالباً لأن خواصه إحصائية، ولأنه دُرس أكثر من القياسات الأخرى، وهو ليس سالباً وتابعاً قابلاً للاشتقاق، ويعاقب الأخطاء الكبيرة أكثر من الأخطاء الصغيرة. ويعود سبب شعبيته لكونه مشابهاً لقياسات إحصائية أخرى مثل ارتباط العينة الذي ينتهي إلى الواحد كلما انخفض الفرق بين نماذج الخرج المحسوب والمنشود.

بعد اعتماد المعادلة (38.6) لتكون القياس المختار للخطأ، يمكننا التساؤل عن كيفية تدريب الشبكة بفعالية أكثر لإيجاد الأصغر الكلي (global minimum). هذا هو السؤال الذي انصب عليه الاهتمام عبر السنين القليلة الماضية، مع تحقيق تحسينات عملية التدريب بالانتشار الخلفي المرتبطة بشكل سطح الخطأ.

1.7.6 سطح الخطأ The error Surface

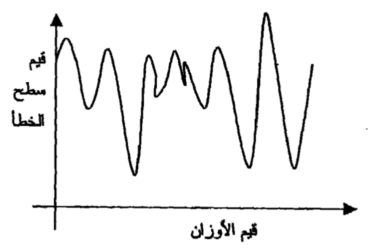
للشبكات العصبونية الصنعية المتعددة الطبقات بتغذية أمامية ذات توابع التفعيل غير الخطية سطحُ خطأ يمثل متوسط مربع الخطأ (MSE) في فراغ الأوزان R^Q ذي البعد Q. إن هذا السطح، بوجه عام، ليس سطحاً قطعياً ناعماً كما في حالة تابع التفعيل الخطي لشبكات الطبقة الواحدة.

على العموم، سيكون سطح الخطأ معقداً، ومن المعتقد أن له قيماً أصغرية كليةً ومحليةً عديدة (McInerny وزملائه عام 1989 [28]) كما هو موضح في الشكل (9.6) لفراغ وزن

ثنائي البعد. تكثر القيم الصغرى جزئياً بسبب تناظرية الأوزان والعقد في الشبكة.

في حالة أية قيمة معطاة للخطأ E (تشمل المناطق الدنيا)، تعطي تعديلات عديدة للأوزان نفس الزيادة لقيمة E. تحدث القيم الصغرى أيضاً نتيجةً لحذف التأثيرات بين الأوزان المتعاكسة الإشارات بالإضافة إلى مجموع عدة مركبات غير خطية عبر فراغ النموذج.

وبالنتيجة، فإن خوارزمية الانتشار الخلفي لن تستطيع التأكيد مطلقاً أنها قادرة على إيجاد الأصغر الكلي كما في حالة قاعدة دلتا لشبكة وحيدة الطبقة.



الشكل 9.6: خطأ نموذجي لشبكات MLFF بتوابع تفعيل غير خطية

على العموم، في حالة الشبكات المتعددة الطبقات الأمامية التغذية، سيكون لأسطح الخطأ قيمٌ دنيا محليةٌ عديدة. والاستثناء الوحيد لهذه القاعدة سيكون في صنف الشبكات MLFF التـــي تحقق شروطَ نظرية kolmogorov (Yu عام 1991[59])؛ التـــي تنص على ما يلي:

في حالة أي تابع مستمر $f(x_1, x_2 ..., x_n)$ لعدة متحولات معرف على (I^n) $n \geq 2$، حيث $I = [0, 1]$، يمكن أن يمثل كما يلي:

$$f(x) = \sum_{j=1}^{2n+1} \chi_j \left(\sum_{i=1}^{n} \varphi_{ij}(x_i) \right) \quad (40.6)$$

حيث χ_j و φ_{ij} هما تابعان مستمران لمتحول واحد و φ_{ij} هي توابع انسيابية لا تعتمد علـــى f. ولهذه الشبكات المحققة لشروط النظرية عدد غير محدود من القيم الدنيا وجميعها قيمٌ دنيا كلية (ليست محلية).

ولكن، حتـــى من أجل هذه الشبكات، من الممكن لعملية التعليم أن تصل إلى حالة

سكون أو التصاق (عدم التقدم أو التراجع)، مثلاً في منطقة مسطحة أو شبه مسطحة من سطح الخطأ.

عند بداية عملية التدريب يبدأ بحث تدرج الهبوط عند محل بقيمة خطأ E معينة بواسطة قيم الأوزان الأولية W(0) وزوج نموذج التدريب $(\mathbf{x}^p, \mathbf{t}^p)$، حيث

$$E = \frac{1}{P}\sum_{p=1}^{P} E^p = \frac{1}{2P}\sum_{p,k}(t_k^p - f_k(\mathbf{W}(0) - \mathbf{x}^p))^2 \qquad (41.6)$$

خلال التدريب، تُعيِّن حسابات تدرج الهبوط كيف يجب أن تعدل الأوزان على نحو متزايد عند كل محل جديد لتتحرك سريعاً في الاتجاه المعاكس لاتجاه الصعود الشديد (أي الهبوط الشديد). بعد إجراء تعديل الأوزان المتزايد، يزاح المحل إلى محل آخر بقيمة E مختلفة على سطح الخطأ ـــ وزن.

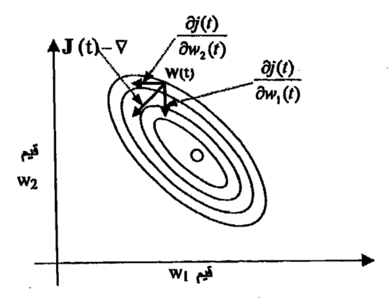
تتكرر هذه العملية لكل نموذج تدريب p (أو كل دور)، ويزاح المحل تتابعياً إلى مستويات الخطأ الأخفض حتـــى الوصول إلى قيمة عتبة الخطأ أو حتـــى الوصول إلى نهاية العدد الكلي المحدد من دورات تكرار التدريب.

في التحرك لأسفل سطح الخطأ ـــ وزن، لن يكون المحل المتبوع عموماً ممراً مثالياً. بل سيعتمد ذلك على شكل السطح وعلى معدل التعليم η الذي سنشرحه فيما بعد. هناك مناطق مسطحة واسعة ومنخفضات حيث يجب أن تعدل الأخطاء مرات عديدة لتحقيق السقوط في الخطأ الأصغري. عند المحلات ذات الانحدارات الشديدة، يمكن أن تنتج خطوات كبيرة في حركات اهتزازية عبر الانحدارات. سيجعل بعض الشذوذ من الصعب على تدرج الهبوط اختيار الاتجاه الصحيح للحركة، وبذلك سيكون هناك تقدم بطيء وغير محدد الاتجاه.

وبسبب كون سطح الخطأ هو مجموع حدود مربعة تصف مقاطع قطعية عوضاً عن دائرية، فإن تدرج الهبوط عادة لن يتجه مباشرة إلى اتجاه الأصغر (Jacobs عام 1988 [60])، وهذا موضح في الشكل (10.6) في حالة سطح ثنائي البعد.

وبسبب أن سطح الخطأ سيكون أشد انحداراً على طول البعد w_2 منه على طول البعد w_1، فإن مشتق w_2 سيكون أكبر من مشتق w_1، وهذا يعطي في المحصلة الشعاعية أن الإزاحة ستكون أكبر في اتجاه مشتق w_2. أي إن المحصلة الشعاعية لن تصوب مباشرة إلى

قيمة مكان الأصغر الصحيحة.



الشكل 10.6: اتجاه الهبوط لحالة ثنائية البعد.

8.6 تحسين معدل التقارب

Improving the rate of convergence

رأينا أن عملية البحث عن الأصغر الكلي تعتمد على شكل سطح الخطأ بالإضافة إلى خوارزمية التعليم ومجموعة التدريب. سنناقش في هذه الفقرة الموضوعين الأخيرين وهما إجراء التعليم ومجموعة التدريب، وسنركز الاهتمام على إيجاد الإجراءات التـــي تحسن معدل تعليم التقارب.

1.8.6 الأوزان الأولية

نذكّر بأن بداية عملية التعليم هي اختيار القيم الأولية للأوزان. إن انتقاء قيم مختارة سلفاً يؤثر في معدل التقارب. وينتج عن اختيار قيم أولية W(0) قريبة من *W تقاربٌ سريع. من ناحية أخرى، يمكن أن يسبب الاختيار السيئ للأوزان الأولية عشرات الآلاف من عمليات التكرار حتـــى الوصول إلى التقارب، وقد لا نصل إلى التقارب مطلقاً.

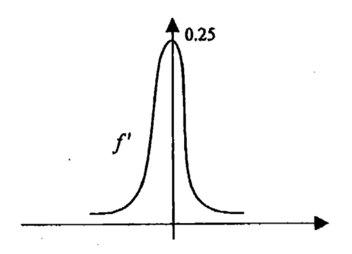
مثلاً، سيؤدي اختيار جميع قيم الأوزان بنفس القيمة إلى عدم الوصول إلى التقارب لأن كل وحدات الطبقة المخفية المتصلة مع وحدات الخرج ستستقبل قيم تحديث خطأ متساوية (لأن الخطأ المنتشر عكسياً متناسب مع الأوزان). ولما كان تعديل الأوزان يعتمد على

إشارات الخطأ المنتشرة عكسياً، فإن كل تعديلات الأوزان ستكون متساوية، وسيبقى النظام ساكناً عند نقطة التوازن التـــي ستمنع الأوزان من التغير.

من الواضح أنه يمكن أن تسبب اختيارات أخرى لقيم الأوزان الأولية وصول عملية التعليم إلى حالة الالتصاق والسكون في منطقة مسطحة على سطح الخطأ- وزن. فقد ثبت أن وضع قيم صغيرة للأوزان الأولية (ولكن ليست صغيرة جداً) كيفياً، وليكن بين +1 و-1، يعتبر طريقة فعالة لتجنب المنخفضات الضحلة وإمكانية الوقوع في العثرات عند بداية عملية التدريب. إما إذا اختيرت الأوزان بقيم كبيرة، فإن تابع التفعيل سيصبح مشبعاً. وعندما يحدث هذا فإن مشتق تابع التفعيل في حدٍّ تحديث الطبقة المخفية:

$$\upsilon_{ji}^{new} = \upsilon_{ji}^{old} + \Delta\upsilon_{ji} = \upsilon_{ji}^{old} + \eta x_i f'(H_j)\sum_k \delta_k w_{kj} \qquad (42.6)$$

سيكون قريباً من الصفر. وذلك لأن مشتق تابع sigmoid هو $f(x)\,[1-f(x)]$ حيث $f(x)$ هو خرج أي وحدة sigmoid. يأخذ هذا المقدار قيمته العظمـــى فـــي النقطـــة $x = 0$ $(f'(0) = 0.25, \beta=1)$ وتقترب من الصفر كلما تحرك $f(x)$ باتجاه الصفر أو الواحد. لذا تجري عند هذه النقاط، تعديلات صغيرة للأوزان. وسيكون التقدم إلى المنحدر بطيئاً، كما هو موضح في الشكل (11.6).



الشكل 11.6: تعديل الأوزان متناسب مع متشق تابع التفعيل f'

إن إحدى الطرائق لتجنب هذا الإشباع هي استعمال قياس الخطأ الذي ينمو كثيراً كلما تحرك$f(x)$ باتجاه الصفر أو الواحد. وهناك طريقة أخرى بسيطة، تتلخص بإضافة ثابت b إلى

حدٍّ المشتق لكي لا ينقص مطلقاً إلى ما دون هذه القيمة. أثبتت تجارب Fahlman عام 1988 [61] أن استعمال قيمة b = 0.1 ستعطي تحسينات كافية في معدل التقارب عبر طرائق الانتشار الخلفي التقليدية.

في تجارب المحاكاة الأخرى التـــي أجراها Hirose عام 1991 [62]، أثبت أن التقارب يمكن ألاّ يدرك في بعض الشبكات عندما تختار القيم الأولية للأوزان اختياراً سيئاً.

لقد وجد من بعض التجارب التـــي أجريت على شبكات MLFF بطبقتين لتنفيذ التابع XOR (الشبكة مؤلفة من طبقة واحدة مخفية بعقدتين وعقدة وحيدة في الخرج وهي تشبه شبكة مادالينMADALINE للتابع XOR) ومسائل تعرّف الأشكال 16× 16 عنصر صورة، التـــي كان فيها عدد الوحدات المخفية متغيراً لدراسة خواص تقارب الشبكات، أن القيم الأولية الكيفية للأوزان الصغيرة جداً أو الكبيرة جداً أدت إلى إخفاق التقارب.

من ناحية أخرى، عندما عينت القيم في المجال [0.5 ,0.5 –] أو [1, 1–] جرى الوصول إلى التقارب على نحو دائم. مثلاً في حالة التابع XOR المنفذ بشبكة MLFF وباستعمال خوارزمية الانتشار الخلفي، لوحظ أن قيم الأوزان الصغيرة جداً (ضمن المجال [0.05 ,0.05 –]) لا تحقق التقارب مطلقاً، وفي هذه الحالة تقع الحسابات في مشكلة الأصغر المحلي. أما في حالة قيم الأوزان الأولية في المجالات التالية:

[(–0.25, 0.25), (–0.5, 0.5), (–1, 1), (–1.5, 1.5), (–2.5, 2.5), (–5, 5)] فقد كان التقارب محققاً مع ملاحظة ازدياد عدد التكرارات بزيادة قيم المجالات طردياً وهذا ما أدى إلى زيادة زمن التعليم.

2.8.6 تقدير القيم الأولية Estimating initial values

بدلاً من تخصيص قيم أولية للأوزان، يمكن تقدير القيم التـــي تكون قريبةً لحلٍّ أصغري وأمثلي للشبكة، ومن ثم تؤخذ هذه القيم لتمثل القيم الأولية للأوزان W(0) عند الانطلاقة الأولى للخوارزمية، بدلاً من وضع قيم على نحو عشوائي (Chen & Nutter عام 1991[63]). ولمعرفة ذلك سنستعرض بعض التعاريف:

x: مصفوفة نماذج دخل التدريب ببعد P × n، يمثل كل شعاع سطر فيها نموذج دخل

التدريب في R^n.

T: مصفوفة نماذج الخرج المنشود ببعد $P \times m$، يمثل كل شعاع سطر نموذج الخرج في R^m.

H: مصفوفة مخارج الوحدات المخفية ببعد $P \times h$ لـ P نموذج تدريب.

F: تابع شعاع قابل للقلب، لكل مركبة تابع شعاع عمود مقلوب f^{-1} .

V: مصفوفة وزن الطبقة المخفية ببعد $n \times h$.

W: مصفوفة وزن طبقة الخرج ببعد $h \times m$.

لاحظ أنه في حالة تابع sigmoid: $f(x)=[1+\exp(-x)]^{-1}$ ويعطى مقلوب هذا التابع لكل مركبة fمن f كما يلي: $f^{-1}(x)=-\ln[1/x-1]$. لدينا أيضاً العلاقات التالية:

$$f(\mathbf{X}\,\mathbf{V})=\mathbf{H}\ ,\ f(\mathbf{H}\,\mathbf{W})=\mathbf{T} \qquad (43.6)$$

لتقدير قيم أوزان حل الشبكة، سنبدأ بإعطاء الأوزان القيم البدائية V وهي قيم كيفية في المجال $[-\lambda, \lambda]$، حيث $\lambda > 0$ قيمة صغيرة. ومن ثم يجري حساب القيمة **H** من (41.6). ولإيجاد تقدير الأوزان **W** نستعمل العلاقة: $f^{1}(\mathbf{T})=\mathbf{H}\,\mathbf{W}$.

هناك ثلاث حالات تتعلق بالاعتماد على الرتب r (rank) للمصفوفات **H** و$\mathbf{H}\,f^{1}(\mathbf{T})$.

1. في حالة كون:

$$r(\mathbf{H})=r(\mathbf{H}\,f^{1}(\mathbf{T}))=h \qquad (44.6)$$

يوجد حل وحيد يعطى بـــ:

$$\mathbf{W}=\mathbf{H}^{-1}\,f^{1}(\mathbf{T})$$

2. وفي حالة كون:

$$r(\mathbf{H})=r(\mathbf{H}\,f^{1}(\mathbf{T}))<h \qquad (45.6)$$

هناك عدد لانهائي من الحلول. في هذه الحالة ليس هناك طرق تحليلية لتحديد الحل الأمثل، لذا يمكن استعمال حل النظيم الأصغر لإعطاء:

$$\mathbf{W}=\mathbf{H}^{T}(\mathbf{H}\,\mathbf{H}^{T})^{-1}f^{1}(\mathbf{T})$$

حيث $\mathbf{H}^{T}$ هو منقول **H** و$\mathbf{H}^{T}(\mathbf{H}\,\mathbf{H}^{T})^{-1}$ هو المقلوب الوهمي لـــ **H**.

3. أخيراً إذا كان:

$$r(\mathbf{H}) < r(\mathbf{H}\, f^{1}(\mathbf{T})) \tag{46.6}$$

عندها لا يوجد حل دقيق، لذا يمكن إهمال n-m نموذج تدريب لجعل المعادلة (46.6) بحالة محددة وحيدة، أو إيجاد حل خطأ المربعات الصغرى بحيث يكون E(W) أصغرياً. في الحالة الأخيرة يمكن استعمال المقلوب الوهمي مرة ثانية لإيجاد W من العلاقة:

$$\mathbf{W} = (\mathbf{H}^{T}\mathbf{H})^{-1}\mathbf{H}^{T} f^{1}(\mathbf{T}) \tag{47.6}$$

بعدئذ يمكن تطبيق خوارزمية الانتشار الخلفي لتعديل الأوزان V وW.

وبالتبادل، يمكننا استعمال تقديرات W الناتجة فيما سبق لتنفيذ تقدير خلفي لـ V بأسلوب مماثل لإيجاد تقديرات W في خوارزمية التقدير الأمامي. وبتطبيق تقديرات أمامي ـ خلفي تكرارياً يمكننا عندئذ الحصول على تقدير تكراري لكلا V وW.

لقد استنتج من تنفيذ ست تجارب على شبكات MLFF صغيرة أن توفير زمن الحسابات وصل إلى 50% باستعمال هذه التقنية مقارنة مع طرق وضع قيم الأوزان الأولية كيفياً.

أثبتت طريقة أخرى لتعيين قيم الأوزان الأولية أنه يمكن توفير زمن حسابات (في حالة شبكات MLFF المستعملة كمصنِّفات) وذلك بوضع الأوزان بين طبقة الدخل والطبقة المخفية الأولى v_{ij} بقيم نماذج بدائية معيارية P_j (أشعة صف مرجعية)، والأوزان بين الطبقة المخفية وطبقة الخرج w_{jk} مساوية للواحد إذا كان P_j شعاع بدئي يمثل الصف رقم k وإلا سيساوي الصفر (Denoeux وزملاؤه عام 1991[64]).

إن النماذج البدائية هي نماذج مجموعة التعليم وهي ممثلات جيدة للصفوف. مثلاً، يمكن أن تكون مراكز متوسطة للتجمعات، أو نماذج الرموز المُكمَّمة شعاعياً، أو أشعة نوع القيمة المتوسطة المعينة بواسطة تقنيات تشكيل التجمعات المعروفة جيداً. لذا ستكون أشعة التدريب البدائية معيارية بطول يساوي الواحد.

اقترح Kim & Ra عام 1991 [65] وHaario & Jokinen عام 1991 [66] طرقاً أخرى لتقدير الأوزان الأولية. مثلاً، إذا أصبح الوزن w_i معيارياً كما يلي:

$$u_i = \frac{w_i}{\sqrt{\sum_{i=1}^{n} w_i^2}} \tag{48.6}$$

فيمكن إثبات أنه لكي يكون u متقارباً، سيكون لدينا:

$$0 < \frac{w_i}{\sqrt{\sum_{i=1}^{n} w_i^2}} < 1 \qquad (49.6)$$

ومن ثَم، إذا افترضنا أن مطالات القيم الأولية متساوية تقريباً، فإن قيم الأوزان الأولية $w_i(0)$ ستحقق:

$$|w_i(0)| > \sqrt{n/N} \quad , \quad i = 1, 2, ..., N \qquad (50.6)$$

لوحظ توفير معتبر في زمن التعليم مقارنة مع الانتشار الخلفي التقليدي (بحد كمية حركة) في مسائل XOR والتكافؤ ثلاثي الخانة.

3.8.6 تقديم ضجيج عشوائي

يمكن تجنب مشكلة الأصغر المحلي بتقديم ضجيج عشوائي خلال عملية البحث. يعتبر إعطاء ضجيج عشوائي طريقة أخرى لإزاحة محل تابع الخطأ والسماح بذلك لعملية الانحدار بالهروب من الأصغر المحلي والمتابعة نزولاً للبحث عن الأصغر الكلي.

تنجز عملية تقديم الضجيج بإضافة اضطرابات عشوائية صغيرة إلى الأوزان عندما يتوقف التقارب، وذلك بإضافة ضجيج ما مباشرة إلى النماذج، أو بواسطة اختيار كل نموذج مقدم بعملية عشوائية أو بواسطة عملية إحصائية أخرى مثل محاكاة التلدين (سنشرح ذلك في الفصل التاسع).

يمكن أن يحسِّن تقديمُ الضجيج خلال عملية التعليم مقدرةَ الشبكة على التعميم كما سنرى فيما بعد، ولكن الضجيج يمكن أن يسبب مشاكل أيضاً. مثلاً، من الممكن ضياع الأصغر المحلي الجيد إذا غيّر الضجيج على نحو كافٍ المكان على سطح الخطأ خلال الهبوط.

4.8.6 تأثيرات عامل معدل التعليم

يعيِّن عاملُ معدل التعليم η في قاعدة دلتا المعممة حجمَ تعديلات الأوزان عند كل تكرار، ومن ثَم فهو يؤثر في معدل التقارب. تعتبر قيمة η هامةً أنه يمكن أن تنتج تغيرات كبيرة في معدل التعليم من اختيارات مختلفة لـــقيمة η. ينتج عن الاختيار السيء إخفاق التقارب نهائياً.

من المعروف أنه من أجل الحصول على نتائج جيدة فإن η لن تكون له قيمة ثابتة خلال عملية التعليم. وهذا ناتج عن أن أسطح الخطأ التـــي لها أشكال مسطحة واسعة قرب الصغريات (أماكن الصغريات المحلية) تتطلب قيم عامل معدل تعليم أكبر من أجل التقارب السريع، على حين أن الأشكال الضيقة الشديدة الانحدار قرب الصغريات تتطلب قيماً صغيرةً لـــ η لتجنب تجاوز أو تخطي الحل.

إذا اختيرت قيمة η كبيرة جداً في حالة سطح الخطأ، فإن ممر البحث سيهتز حول الممر المثالي وسيتقارب ببطء أكثر من الهبوط المباشر، ويمكن أن يبتعد. من ناحية أخرى، إذا اختيرت قيمة η صغيرة جداً، فإن الهبوط سيتقدم في خطى صغيرة جداً وهذا سيزيد الزمن الكلي للوصول إلى التقارب.

يوضح الشكل (12.6) ممرات التقارب النموذجية لقيم مختلفة لـــ η استعملت في تعليم الترافقات العشوائية ، يبين هذا الشكل أن قيمة $\eta = 0.9$ هي قيمة مثلى.

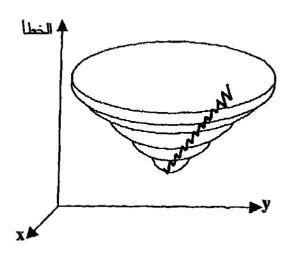
بالطبع، لا تكون عادةً سطوح الخطأ للشبكات **MLFF** غير الخطية قطوعاً ناعمةً كما هو موضح في الشكل (12.6).

على أية حال، الاختيار الأفضل لـــ η سيكون مسألة اعتمادية، وقد تتطلب بعض التجارب قبل أن يوجد الاختيار الصحيح. وللحصول على أفضل النتائج، نستعمل قيماً مختلفةً لـــ η لكل وزن خلال عملية التدريب وذلك بسبب عدم وجود قيمة وحيدة مثالية لكل الأبعاد. وكان هذا جلياً في الشكل (10.6)، إذ لوحظ أن مشتقات الخطأ لكل وزن ستختلف ما لم يكن سطح الخطأ دائرياً تماماً في الشكل.

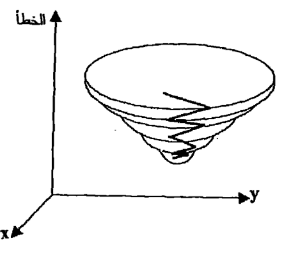
وإذا كان لكل وزن قيمة η_i فسيكون مسموحاً لها التغير مع الزمن. يمكن أن ينجز هذا بالمراقبة المستمرة للتغير في الخطأ خلال التدريب وتعديل قيمة η_i لتكون أكثر ملاءمة لمنطقة الهبوط المحلي. فمثلاً، في حالة قيمة η_i ثابتة، إذا انخفض الخطأ في عدة تعديلات للوزن، فمن المحتمل أن تزداد η_i وذلك بسبب أن قيمتها الحالية محافظة جداً فيما يتعلق بشكل السطح المحلي. من ناحية أخرى، إذا لم تنتج زيادة مع التعديل، فهذا يعنـــي أن العملية تجاوزت الأصغر، ومن ثَم ستنخفض قيمة η_i.

اقترح بعض الباحثين (Jacobs عام 1988[60]) أن يكون لـــ η_i قيمة متكيفة، وفي هذه

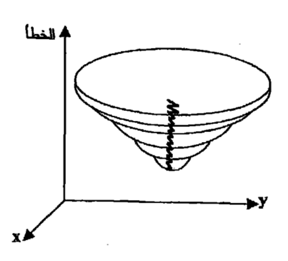
الحالة ستكون قيمة η_i تابعاً لمشتقات الخطأ على التحديثات المتتابعة؛ فهي تُعدَّل قبل كل تحديث وزن خلال عملية التعليم. وقد أثبتت تجارب Tollenaere عام 1990[67] وHirose عام 1991 [62] على شبكة ترافق ذاتـــي بحجم 10 × 10 × 10 (عدد وحدات الدخل × عدد الوحدات المخفية × عدد وحدات الخرج) وشبكة ترافق نموذج عشوائي بحجم 10 × 5 × 2 كيف يمكن أن تؤثر قيمة η في معدل التعليم في شبكات MLFF عندما استعملت لتعلم مهام ترافق نموذج دخل/ خرج.



الشكل (12.6 أ) ممر التقارب في حالة $\eta = 0.2$ (هبوط بطيء بخطأ صغيرة)



الشكل (12.6ب) ممر التقارب في حالة $\eta = 0.9$ (هبوط سريع بخطأ كبيرة)



الشكل (12.6.ج) ممر التقارب في حالة η = 1.5 (هبوط بطيء بخطأ صغيرة)

لوحـــظ مـــــن الـــتجارب علـــى الشبكة الأولى أن الزيادة فـــي قيمة η (ضمن المجال [0.00 – 0.40]) تعطـــي نقصاناً أسياً في معدل التعليم حتـــى نقطة معينة وهذا مفضل، وأن أزمـــنة التعلـــيم في حالة قيم η ضمن المجال [0.15 – 0.30] ستكون صغيرة جداً وتبدأ هذه الأزمـــنة بالازديـــاد الكبير خارج هذا المجال، وخاصة عند قيمة η أصغر من0.05 وأكبر من 0.35.

أما من التجارب على الشبكة الثانية فقد وُجد أيضاً أن الزيادة في قيمة η (ضمن المجال [0.00 – 2.00]) تعطي نقصاناً أسيّاً في معدل التعليم حتـــى نقطة معينة، وأن أزمنة التعليم في حالة قيم η ضمن المجال [0.2 – 1.3] ستكون صغيرة جداً وتبدأ هذه الأزمنة بالازدياد الكبير خارج هذا المجال، وخاصة عند قيمة η أصغر من 0.15 وأكبر من 1.75.

5.8.6 إضافة حد كمية الحركة

هناك طريقة أخرى لتحسين معدل التقارب تتلخص بإضافة عطالة ما أو كمية حركة إلى تعبير التدرج. يمكن أن ينجز هذا بإضافة جزء تغير الوزن السابق إلى تغير الوزن الجاري. يمكن أن تُنعِّم هذه الإضافة ممر الهبوط السابق بمنع تغير كبير في التدرج وفقاً للشذوذات المحلية. ويمكن استخدام حد كمية الحركة كمعدل تأثير لتنعيم مسار التدرج كلما تحرك إلى أسفل المنحدر.

تتضمن قاعدة التحديث العامة المقدمة من قبل Rumelhart عام 1986 [56] حداً لكمية

الحركة، ومعادلة التحديث التـــي استعملت هي:

$$\Delta w_j(t+1) = -\eta \frac{\partial E}{\partial w_j(t)} + \alpha \Delta w_j(t) \qquad (51.6)$$

حيث α هو عامل كمية الحركة وقيمته موجبة وأقل من الواحد. وتقع قيمة هذا العامل النموذجية ضمن المجال [0.9 – 0.5] ولكن في بعض التطبيقات تكون قيمة $\alpha = 0$ هي أفضل القيم (Fahlman عام 1988 [61]).

من الملاحظ أن المعادلة (51.6) هي معادلة فروق، أي إن حد كمية الحركة هو فعلاً مجموع مثقل أسي للمشتقات الجزئية للأوزان الحالية والماضية. يشجع هذا الحد الحركة في نفس الاتجاه بخطى متتابعة، لذا فإنه سيضغط أية اهتزازات يمكن أن تنتج من التغيرات في انحدار سطح الخطأ.

من الأفضل تعديل ثابت كمية الحركة ديناميكياً بالإضافة إلى ثابت التعليم. ولما كانت القيمتان ليستا مستقلتين إحداهما عن الأخرى، فستختار قيمة لكلا الثابتين ويجري تعديلهما كزوج.

نفذت تجارب عديدة من قبل الباحثين لتعيين تأثيرات حد كمية الحركة على سرعة التقارب واستقرارية إجـــراءات التعليم أمثال Fahlman عام 1988 [61]، وJacobs عام 1988 [60]، وMinai & Williams عـــام 1990 [68]، وTollenaere عـــام 1990 [67]، وSato عام 1991 [69].

ومن نتائج تجارب Sato عام 1991 على مرمز 2.4 ومسألة التكافؤ لـــ 4 خانات، حيث تغيرت قيمة α من 0.0 حتـــى 0.9 بخطوة 0.1 ومن 0.9 حتـــى 0.99 بخطوة 0.01 على الترتيب، يمكن استنتاج العلاقات التالية (صحيحة فقط في حالة عدد ضخم من الأدوار):

في حالة مسألة المرمز $\qquad N = 8.28 \times 10^2 \propto (1-\alpha)$

في حالة مسألة التكافؤ $\qquad N = 3.47 \times 10^4 \propto (1-\alpha)$

حيث N عدد أدوار التعليم حتـــى بلوغ التقارب.

ونلاحظ أنه إذا كانت قيمة α صغيرة فإن $N \propto (1-\alpha)/\eta$. هذا يقترح أن α يتصرف كمكامل بدلاً من مرشح تمرير منخفض كما طلب من قبل تجارب أخرى. عندما يكون α

كبيراً فإن العلاقة لن تطبق جيداً كحد كمية حركة، ومن ثم يزيد مسار Δw في فراغ الوزن. وهذا سيسرع أو سيبطئ التعليم على حسب سطح طاقة ـــ وزن.

6.8.6 تأثيرات توابع التفعيل على التقارب

من الكتب والمقالات المنشورة في حقل الشبكات العصبونية الصنعية نرى أن أكثر توابع التفعيل استخداماً للشبكات العصبونية المتعددة الطبقات الأمامية التغذية هي توابع التفعيل sigmoid المشروحة سابقاً.

ومع أن هذا النوع من التوابع معقول بيولوجياً وله بعض الخواص الحسابية المفضلة، فإنه يسبب بعض المشاكل الخطيرة من وجهة نظر التقارب:

1. يعتقد أن لسطح الخطأ المتشكل من خلال استعمال توابع التفعيل sigmoid شذوذاً كبيراً (الموصوف سابقاً) يؤدي إلى تأخير التقارب أو حتـــى قد يمنع التقارب.
2. عندما تقترب قيمة خرج العصبون من الصفر أو الواحد، فإن العوامل $O_k(1 - O_k)$ في خوارزمية تحديث الانتشار الخلفي (المعادلة (29.6)) ستجعل إشارات الخطأ صغيرة جداً. وحتـــى وحدات الخرج بأخطاء كبيرة لن تنتج تعديلات تحديث قوية، وبذلك يصبح التقارب بطيئاً. يمكن أن تحدث نفس الظاهرة عندما تصبح الأوزان الليفية للعصبون كبيرة، إذ أنها تجبر العصبون على الإشباع.
3. ثبت أن عدم تناظر تابع sigmoid القياسي سيجعله أدنى من تابع sigmoid المتناظر (الظل القطعي) من وجهة نظر التقارب (Fogelman Soulie عام 1991 [70]).

يظهر من جميع هذه المحددات أنه من المستغرب أن يكون لتابع sigmoid هذه الشعبية مستقبلاً. بالطبع ليست سرعة التقارب وكلفة الحسابات هي الاعتبارات الوحيدة التـــي تؤخذ بعين الاعتبار عندما نختار توابع التفعيل. فهناك اعتبارات هامة أيضاً مثل مقدرات التطبيق والتعميم التـــي تعتمد هي بدورها على اختيار توابع التفعيل.

7.8.6 تأثيرات تابع الخطأ في التقارب

لاحظنا من قبل أن تابع الخطأ الممثل بمتوسط مربع الخطأ (MSE) لم يكن الاختيار الأفضل لقياس إنجاز أو كلفة الشبكات MLFF. إن سطح الخطأ لمتوسط مربع الخطأ له شواذ

عديدة تجعل التقارب صعباً.

تبدي بعض البدائل الممكنة لمتـــوسط مربع الخطأ مثل توابع الأساس اللوغارتمي للخطأ (z_k-t_k) آمالاً كبيرة. فقد اقتُرح تابعان اثنان سنصفهما كما يلي: الأول اقترحه Matsuoka & Yi عام 1991 [71]، حيث أكد أن سطح الخطأ لتابع الخطأ اللوغارتمي مشابه لأنتروبـــي نظرية المعلومات إذ له صغريات محلية أقل من تابع خطأ MSE. يعرف التابع اللوغارتمي هذا كما يلي:

$$L_1(\mathbf{W}) = \sum_p \left(t_k^p \ln\left(\frac{t_k^p}{z_k^p}\right) + (1-t_k^p)\ln\left(\frac{1-t_k^p}{1-z_k^p}\right)\right) \qquad (52.6)$$

حيث تنحصر قيم t_k^p و z_k^p في المجال [0, 1] وسنفترض $0^*\ln 0 = 0$. عندما استعمل تابع الكلــفة هذا بالانتشار الخلفـــي، فإن الخطأ المنتشر عكسياً مـــن وحدة الخرج k سيساوي $(z_k^p - t_k^p)$ مقارنة مع $z_k^p(1-z_k^p)(z_k^p - t_k^p)$ في حالة MSE. أما الخطأ المنتشر خلفياً من الوحدات الأخرى غير وحدات الخرج فيكون نفسه كما في حالة MSE. وهكذا سيميل التعليم ليصبح أسرع باستعمال L_1(W) بدلاً من E(W) ولن يكون زمن الحساب أكبر.

نفذت مقارنة أزمنة التعليم في مسألة التكافؤ لأربع خانات باستعمال كلا التابعين وتدريب الانتشار الخلفي من قبل Matsuoka & Yi عام 1991 [71]، وقد حصل على نتائج مشابهة في مسائل أخرى .

لوحظ من هذه النتائج أنه في حالة MSE يظهر التعليم أنه وقع في مشكلة الأصغر المحلي من أجل أربع اختبارات من أصل خمسة، على حين تقاربت كل الاختبارات باستعمال توابع الخطأ اللوغارتمية. وهكذا نرى أنه على الرغم من شعبية MSE الواسعة، فإن استعمالها في قياس الخطأ أمر قابل للتساؤل.

اقترح Van Ooyen & nienhuis التابع اللوغارتمي الثانـــي عام 1992 [72] كما يلي:

$$L_2(\mathbf{W}) = -\sum_{p,k}\left[t_k^p \ln z_k^p + (1-t_k^p)\ln(1-z_k^p)\right] \qquad (53.6)$$

تذكـــر أن $0 < z_k^p < 1$. لوحـــظ أن هـــذا التابع أيضاً له شكل مشابه لأنتروبـــي نظرية المعلومات. ولدينا في هذه الحالة :

$$\frac{\partial L_2}{\partial z_j} = \frac{z_j - t_j}{z_j(1-z_j)}$$
$$\frac{\partial L_2}{\partial w_{ji}} = z_j - t_j \qquad (54.6)$$

وهكذا (في حالة L2) يكون الخطأ المنتشر عكسياً من كل خرج متناسباً مع الفرق بين الخرج المحسوب والخرج المنشود. لقد أثبتت تجارب محاكاة باستعمال L2 تحسن معدل التقارب نوعاً ما مقارنة مع MSE. وكذلك فإن التحسينات كانت واضحة أكثر في مسائل أصعب حيث حققت سرعة 2 إلى 1. أعطى تحليل المسافة بين المخارج المحسوبة والمنشودة ، بعد P دوراً أو مسحاً لمجموعة التدريب ككل، التقريبَ التالي:

$$|\mathbf{z}-\mathbf{t}| \approx \frac{1}{P^q} \qquad (55.6)$$

في حالة خوارزمية الانتشار الخلفي الأساسية q = 0.5 وL2، q = 1، ومنه يثبت أن التقارب النهائي أسرع في حالة L2 منه في حالة MSE.

9.6 طرق تعليم أخرى

لقد نُفِّذَتْ محاولات عديدة لزيادة معدل التعليم للانتشار الخلفي؛ فقد وُصف من قبلُ عددٌ من التعديلات والتغيرات التـــي أجريت على طريقة الانتشار الخلفي الأساسية. في هذه الفقرة سنصف بعض الإجراءات الهامة التـــي تختلف نوعاً ما عن طرائق تدرج الهبوط المشروحة سابقاً.

1.9.6 طريقة الانتشار السريع Quickprop

إن إجرائية الانتشار السريع هي فعلياً طريقة المرتبة الثانية التـــي تفرض أن شكل سطح الخطأ لكل وزن هو سطح صاعد؛ فهو يتطلب أن نحتفظ بنسخ مشتق الخطأ $\partial E/\partial w(t-1)$ لكل وزن والفرق بين القيمة الحالية والسابقة للأوزان من دور التدريب السابق. يعطى حساب التحديث في الانتشار السريع كما يلي:

$$\Delta w(t) = \frac{s(t)}{s(t-1)-s(t)} \Delta w(t-1) \qquad (56.6)$$

حيث s(t) و(s(t −1 هما القيم السابقة والحالية للمشتق $\partial E/\partial w$ على الترتيب. أثبتت تجارب على مسائل مختلفة أن نتائج الانتشار السريع أفضل بعامل 4 عن النسخ المعدلة الأخرى للانتشار الخلفي. وكذلك تبدي هذه الطريقة أنه من الممكن تكبير المقياس في حالة مسائل أعقد من نفس النوع.

2.9.6 طريقة Delta-Bar-Delta

Delta-Bar-Delta هي الطريقة التـــي تؤدي أربع عمليات تنقيب أساسية فيما يتصل بتدرج الهبوط. طورً Jacobs هذه الطريقة عام 1988 [60]. وهي تتألف من قاعدة تحديث الوزن وقاعدة تحديث التعليم.

طبقت قاعدة تحديث الوزن على كل وزن $w(t)$ عند التكرار t من خلال العلاقة المعطاة بـــ :

$$w(t+1) = w(t) - \eta(t+1)\frac{\partial E(t)}{\partial w(t)} \qquad (57.6)$$

حيث $\eta(t)$ معدل التعليم لوزن معطى عند تحديث التكرار t.

تُعرَّف قاعدة تحديث معدل التعليم لوزن معطى w كما يلي:

$$\Delta\eta(t) = \begin{cases} \kappa & \bar{\delta}(t-1)\delta(t) > 0 \\ -\varphi\eta(t) & \bar{\delta}(t-1)\delta(t) < 0 \\ 0 & otherwise \end{cases} \qquad (58.6)$$

حيث $\delta(t) = \frac{\partial E(t)}{\partial w(t)}$ هو المشتق الجزئي للخطأ بالنسبة لـــ w عند الزمن t، و

$$\bar{\delta}(t) = (1-\theta)\delta(t) + \theta\bar{\delta}(t-1) \qquad (59.6)$$

حيث κ و ϕ ثوابت استعملت لزيادة أو نقصان معدل التعليم على الترتيب، و$0 < \theta < 1$ هو ثابت أساس النعومة الأسي للزمن t.

تزيد هذه القاعدة معدلات التعليم خطياً وتنقصها أسياً. في هذه الطريقة تكون معدلات التعليم ممنوعة من أن تصبح كبيرة جداً بسرعة عالية جداً. وتؤكد أن الأوزان دائماً موجبة وتتناقص بسرعة .

كما ذكر من قبل، تؤدي قاعدة Delta-Bar-Delta أربع عمليات تنقيب لتحسين معدل التقارب. هي:

1. لكل وسيط(وزن) معدل تعليم فردي خاص به.
2. يسمح لكل معدل تعليم بالتغير عبر الزمن ليعدل التغيرات في سطح الخطأ.
3. عندما يكون لمشتق الخطأ بالنسبة لوزن معين الإشارة نفسها لعدة خطوات تحديث متتابعة، يزداد معدل التعليم لذلك الوزن. وهذا بسبب أن سطح الخطأ يُحتمل أن يكون له انحناء صغير عند بعض النقاط ويستمر الانحدار بنفس المعدل من أجل مسافة ما. لذا سيزداد حجم الخطوة ليسرّع حركة الانحدار.
4. عندما تتناوب إشارة المشتق بالنسبة للوزن في عدة خطوات متتابعة، يخفض معدل التعليم في ذلك الوسيط. وهذا بسبب كون سطح الخطأ له انحناء كبير عند النقطة الحالية والانحدار يمكن أن يغير إشارته بسرعة. وهكذا، ستخفض قيمة الخطوة لمنع الاهتزازات.

أخيراً، لابد من القول بأن طرق المرتبة الثانية تحقق تقارباً أسرع من الطرق المبنية على المرتبة الأولى. ولكن لسوء الحظ، تتطلب هذه الطرق حساب مصفوفة Hessian، وهذا الحساب معقد ومكلف.

لقد قُدِّم عدد من التقريبات التـــي تبسِّط الحسابات مقارنة مع طريقة Hessian ونيوتن وأثبتت فعالية نسبية. فأعطى Battiti عام 1992 [73] مراجعة لهذه الطرق وأعطى Bishop عام 1992[74] الحسابات الكاملة لـــ Hessian لشبكة MLFF وحيدة الطبقة المخفية.

10.6 اختيار عدد الطبقات وعدد الوحدات

لما كان تدريب الانتشار الخلفي مكلف جداً ومرهق، فقد تكون لدينا الرغبة في تخمين العدد الأصغري لأمثلة التدريب اللازمة لبنية شبكة معطاة وبمستوى مرغوب به من الأداء.

التقريب الأمثل والأفضل هو التجربة والخطأ. وبهذا التقريب، سيكون أول اختيار هو البنية التـــي ستقام عليها التجربة، وبأخذ مجموعة تدريب ضخمة على نحوٍ متتالٍ واختبار أداء الشبكة بعد كل فصل تدريب، ستستمر هذه العملية مادام الأداء في ازدياد. وستتوقف حالاً عندما يبدأ الأداء بالتناقص.

يمكن أن توسع عملية التجربة والخطأ لتعيين العدد الأفضل للطبقات المخفية وللعقد في كل طبقة. بالطبع سيزداد زمن الحساب اللازم أسيّاً مع عدد الوسطاء الحرة المسموح في المسألة، ومن ثَم، سنحتاج إلى زمن حساب طويل.

استخدمت الأفكار السابقة في التجارب على الشبكات المدربة بالانتشار الخلفي لإيجاد العدد الأمثل للوحدات المخفية(Hirose عام 1991[62]). مثلاً، يمكن أن تستعمل الخوارزمية للتغير الديناميكي لعدد الوحدات المخفية حتـــى يكون العدد الأصغر الذي يحدث التقارب عنده محققاً.

تبدأ الطريقة بتخصيص عدد صغير من الوحدات المخفية ووضع القيم الأولية لتدريب الانتشار الخلفي التقليدي. خلال التدريب، يفحص الخطأ بعد كل 100 تعديل وزن. إذا كان التناقص في الخطأ أكبر من 1% من قيمته السابقة، يستمر التدريب من أجل 100 تعديل وزن أخرى. وإذا لم يكن التناقص أكبر من 1%، فمن المفترض أن يكون الأصغر المحلي قد وُجد، ويجب أن تضاف عقدة مخفية جديدة بوزن يساوي الصفر .

ستغيّر إضافةُ عقدة جديدة سطحَ الخطأ ـــ وزن، وهكذا يعاد التدريب حتـــى يحصل التقارب أو تضاف عقدة مخفية أخرى. تتكرر هذه العملية حتـــى يحدث التقارب (متوسط مربع الخطأ أقل من 0,01). إذا حدث التقارب قبل تجاوز 10000 تعديل، تزال عقدة ويستمر التدريب.

تتكرر عملية إزالة العقد بهذه الطريقة حتـــى ندرك أن التقارب لا يمكن أن يتحقق في أي زمن، عندها تضاف العقدة النهائية، ويكون عدد العقد المخفية المحقق هو العدد الأصغر للوحدات اللازمة للتقارب.

في بعض التطبيقات، تحققت بهذه الطريقة سرعات أكبر بمرتين إلى ثلاث مرات مقارنة مع الانتشار الخلفي التقليدي. وكذلك، فإن هذه الطريقة توجد العدد الأصغري للوحدات المخفية اللازمة لإنجاز قيمة خطأ أصغري. وبالطبع، فإن هذا العدد قد يكون (أو لا يكون) العدد الأفضل من الوحدات لتعميم أفضل. لذا، يلزم إجراء اختبارات أخرى لتعيين ذلك.

تؤكد التجربة النظرية على مقدرة الشبكة على التعميم. وتتعلق هذه المقدرة بعدد الطبقات المخفية، وعدد الوحدات في هذه الطبقات المخفية، بالإضافة إلى مجموعة التدريب

ونظام التدريب.

لاختبار مقدرة التعميم ، عدلت الشبكات من خلال إزالة عقد داخلية غير أساسية، وحشْر طبقات متطلبة جديدة بعقد مدربة مسبقاً. وقد وجد أنه لا يمكن، بوجه عام، أن تتعلم الشبكةُ مجموعةَ التدريب جيداً إذا كان لها عقد مخفية قليلة جداً. من ناحية أخرى، تميل الشبكات التـــي لها عدد كبير جداً من الوحدات المخفية إلى تذكر مجموعة التدريب ولا تعمم جيداً. لذا، فإن من الأفضل في تطبيقات عديدة استعمال طبقة مخفية بحجم أصغري للمهمة المتطلبة. ولن يكون لإزالة الوحدات الفائضة وغير النافعة تأثير خطير في مقدرة الشبكة على التعميم.

وفي مهام تطبيقات أكثر تعقيداً، يمكن أن تعطي طبقتان مخفيتان إنجاز تعميم (وخطأ) أفضل من الطبقة الواحدة. وهذا بسبب أن الوحدات في الطبقة الواحدة تميل إلى العمل داخلياً على نحوٍ كامل بعضها مع بعض، وهذا يجعل من الصعب تحسين التقريب عند نقطة واحدة دون الإساءة في مكان آخر. وبوجود طبقة مخفية إضافية، تستطيع الوحدات في الطبقة المخفية الأولى تجزئة فراغ الدخل إلى منطقتين، وحساب التوابع ضمن هذه المناطق، على حين تستطيع الوحدات في الطبقة الثانية تركيب هذه المخارج، وحساب التوابع المرغوب بها ضمن المناطق، ووضع الخرج بقيمة صفر في مكان آخر، وهذا يعطي دقة كبرى في التطبيق وتعميماً أفضل.

11.6 تمارين

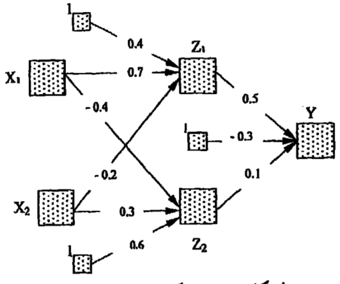
1.6 لتكن لدينا الشبكة MLFF المبينة في الشكل (13.6).

1. أوجد الأوزان الجديدة عندما يقدم للشبكة نموذج الدخل (0, 1) وتكون قيمة نموذج الخرج الموافق مساوية الواحد. باستعمال معدل تعليم $\alpha = 0.25, \sigma = 1$ وتوابع تفعيل sigmoid ثنائية.
2. أوجد الأوزان الجديدة عندما يقدم للشبكة نموذج الدخل (1,1−) وتكون قيمة نموذج الخرج الموافق مساوية الواحد. باستعمال معدل تعليم $\alpha = 0.25, \sigma = 1$ وتوابع تفعيل sigmoid ثنائية القطبية.

3. أوجد الأوزان الجديدة عندما يقدم للشبكة نموذج الدخل (0, 1) وتكون قيمة نموذج الخرج الموافق مساوية 0.8. باستعمال معدل تعليم $\alpha=0.25, \sigma=1$ وتوابع تفعيل sigmoid ثنائية.

4. أوجد الأوزان الجديدة عندما يقدم للشبكة نموذج الدخل (1, 1–) وتكون قيمة نموذج الخرج الموافق مساوية 0.8. باستعمال معدل تعليم $\alpha=0.25, \sigma=1$ وتوابع تفعيل sigmoid ثنائية القطبية.

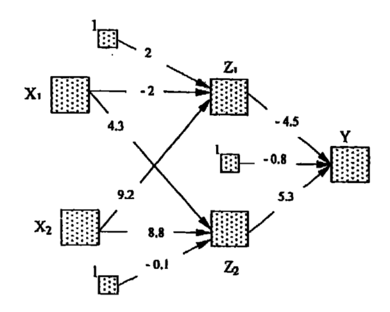
5. كرر الطلبات من 4.1 باستعمال عامل انحدار $\sigma = 3$. هل سيزيد معدل التعليم أو ينقص (حجم تغير الأوزان)؟.



الشكل 13.6: شبكة MLFF للتمرين 1.6

2.6 في الشبكة MLFF، المبينة في الشكل (14.6) المدربة على المعطيات لحل مسألة XOR.

1. باستعمال تابع تفعيل sigmoid ثنائي ، احسب تفعيلات كل وحدة عندما يكون شعاع الدخل (1 ,0). أوجد مقادير δ لوحدات الخرج وللوحدات المخفية. باستعمال $\alpha=0.25, \sigma=1$ احسب تعديل الأوزان، وأوجد الأوزان الجديدة مع الانحياز.
2. كرر الطلب الأول لشعاع دخل (1, 0).
3. علل الفروق بين تغيرات الأوزان على الوصلة إلى وحدة الخرج وتغيرات الأوزان على الوصلة إلى الوحدات المخفية في الطلب الأول والثانـــي.



الشكل 14.6: شبكة MLFF للتمرين 2.6.

3.6 اكتب برنامجاً لأداء شبكة عصبونية مدربة بالانتشار الخلفي وحيدة الطبقة المخفية.استعمل الانحياز على كل وحدة مخفية وكل وحدة خرج ، واستعمل تابع تفعيل sigmoid ثنائي القطبية. ثم اطبع ــ لكل حالة اختبار ــ على الخرج الأوزان الأولية والأوزان النهائية ومعدل التعليم وعدد أدوار التدريب واستجابة الشبكة لكل نموذج دخل عند نهاية التدريب. معطيات التدريب معطاة بالجدول التالي:

XOR ثنائي القطبية

$S(1) = (1, -1),\ t(1) = 1$

$S(2) = (-1, 1)\ ,\ t(2) = 1$

$S(1) = (1, 1),\ t(3) = -1$

$S(2) = (-1, -1),\ t(4) = -1$

استعمل الأوزان الأولية ضمن المجال [0.5 ,0.5–] ومعدلات التعليم التالية:

0.05 و0.25 و0.5. نفذ ــ لكل معدل تعليم ــ 1000 و10000 و25000 دور تدريب (استعمل نفس الأوزان الأولية في كل حالة). استعمل وحدتي دخل ووحدتين مخفيتين ووحدة خرج واحدة.

4.6 اكتب برنامجاً لأداء شبكة متعددة الطبقات مدربة بالانتشار الخلفي وبطبقة مخفية واحدة. استعمل الانحياز على كل وحدة مخفية ووحدة خرج وتوابع التفعيل sigmoid لجميع الوحدات، حيث الشبكة مؤلفة من 10 وحدات دخل و4 وحدات مخفية ووحدتـــي

خرج. نماذج دخل التدريب هي التالي:

نماذج الخرج المنشود	نماذج دخل التدريب
t(1) = −1, −1	s(1) = 0 1 2 3 4 5 6 7 8 9
t(2) = 1, 1	s(2) = 9 8 7 6 5 4 3 2 1 0
t(3) = −1, 1	s(3) = 0 9 1 8 2 7 3 6 4 5
t(4) = 1 , −1	s(4) = 4 5 6 3 2 7 1 8 0 9
t(5) = 1, 1	s(5) = 3 8 2 7 1 6 0 5 9 4
t(6) = 1, −1	s(6) = 1 6 0 7 4 8 3 9 2 5
t(7) = −1, 1	s(7) = 2 1 3 0 4 9 5 8 6 7
t(8) = −1, −1	s(8) = 9 4 0 5 1 6 2 7 3 8

استـــعمل الأوزان الأولية العشوائية الموزعة ضمـــن [−0.5, 0.5] ومعدلات تعليم: 0.05 و0.5، ومن أجل كل معدل تعليم كرر التدريب 5000 و50000 دور (استعمل نفس الأوزان في كل حالة).

5.6 اكتب برنامجاً لأداء وظيفة شبكة عصبونية صنعية مدربة بالانتشار الخلفي لتخزين النماذج التالية:

0#0 #0# ### #0# #0#	##0 #0# ##0 #0# ##0	0## #00 #00 #00 0##	##0 #0# #0# #0# ##0
(-1,-1,-1)	(-1,-1, 1)	(-1,1,-1)	(-1, 1, 1)
### #00 ##0 #00 ###	### #00 ##0 #00 #00	0## #00 #0# #0# 0##	#0# #0# ### #0# #0#
(1, -1, -1)	(1, -1, 1)	(1,1,-1)	(1,1,1)

الشكل 15.6: نماذج دخل التدريب والخرج المنشود المرافق

نماذج الدخل هي أحرف ممثلة بالشكل الثنائي بمصفوفات 3 × 5 ونماذج الخرج المنشود المرافقة هي شعاع ثنائي القطبية بمركبات ثلاثة.

استعمل عدداً متغيراً من الوحدات المخفية (ليس أكثر من 15) ومعدلات التعليم والقيم الأولية للأوزان. قارن نتائجك مع نتائج شبكات الذاكرة الثنائية الاتجاه المعطاة في فصول سابقة .

6.6 اكتب برنامجاً لأداء شبكة عصبوبية مدربة بالانتشار الخلفي بطبقة مخفية واحدة. استعمل انحياز لكل وحدة خرج ووحدة مخفية مع توابع تفعيل sigmoid. دربت الشبكة لتنفيذ التابع التالي:

$$y = f(x_1, x_2) = \sin(2\pi x_1) \cdot \sin(2\pi x_2)$$

في حالة $0 \le x_1 \le 1, 0 \le x_2 \le 1$. سيتغير عدد الوحدات المخفية كجزء من المسألة.

1. حاول أخذ نماذج التـــدريب مفصولة بعضها عن بعض بقيم متساوية (*i*/5, *j*/5) حيـــث *i* = 1, .., 5 و *j* = 1, .., 5، واحسب قيمة الخرج المنشود لكل نقطة (نموذج دخل). اختبر الشبكة على النقاط (*i*/5, *j*/5) حيث *i* = 1, .., 10 و *j* = 1, .., 10. اعرض نتائجك على نحوٍ مشابه لما في المثال الثانـــي المعطى من قبل. هل الاستجابة ستكون أفضل في حالة 10000 دور منها في حالة 1000 دور؟

2. حاوا الإكثار من استعمال النقاط التي تكون قيمة الخرج المنشود فيها غير مساوية الصفر، والإقلال من النقاط التي يكون الخرج المنشود فيها مساوياً للصفر.

3. حاول استخدام نقاط تدريب متولدة عشوائياً.

7.6 اكتب برنامجاً لأداء شبكة عصبونية صنعية متعددة الطبقات أمامية التغذية مدربة بالانتشار الخلفي من أجل ضغط المعطيات في المثال الأول المعطى سابقاً، واستعمل التمثيل الثنائي القطبية للنماذج ونماذج الخرج المنشود مساوية لنموذج الدخل الموافق (ليس من الضروري استعمال 56 وحدة دخل ووحدة خرج).

تطبيقات الشبكات المتعددة الطبقات الأمامية التغذية والانتشار الخلفي

سنصف في هذا الفصل بعض التطبيقات النموذجية للشبكات المتعددة الطبقات الأمامية التغذية. وبسبب الانتشار الكبير والشهرة الواسعة لهذه الشبكات فقد خصصنا فصلاً كاملاً لهذه التطبيقات التـــي تعتبر عينة صغيرة فقط من التطبيقات الكثيرة لهذه الشبكات. فقد أثبتت الشبكات المتعددة الطبقات أنها أدوات تطبيق (mapping) فعالة في مسائل واسعة التنوع، ولذلك استعملت كثيراً من قبل المصممين العمليين في جميع حقول التطبيقات تقريباً. وما على المرء إلا أن يبحث في المؤتمرات السابقة وما تضمنته من مقالات علمية ليرى أنها حفلت بكتب عن التطبيقات المستعملة لهذه الشبكات من الزراعة حتـــى علم الحيوان.

فالمقالات الموجودة في المجلات العلمية والدوريات الأخرى تؤكد عادةً الفوائدَ المكتسبة من استعمال هذه الشبكات مثل التوفير في الكلفة العالية، وتحسينات الدقة، وسرعة الحساب، ..... الخ. وسنناقش هذه الخواص على نحو مستقل في كل تطبيق مدروس فيما يلي.

1.7 تمهيد

إن الأنواع العامة للتطبيقات العملية التـــي استخدمت فيها الشبكات المتعددة الطبقات الأمامية التغذية بنجاح لا تعد ولا تحصى؛ فقد استخدمت في كل التطبيقات المذكورة في الفصل الأول (الفقرة 5.1). واستخدمت هذه الشبكات في حقول عديدة شملت الهندسة، والقانون، وعلوم الحاسوب، وعمليات التحكم، والإحصاء، والتصنيع، والنقل، والمالية،

والاتصالات، والطب،..الخ. يمكن أن تتنافس هذه الشبكات مع البنـــى الأخرى للشبكات العصبونية في معظم مساحات التطبيق باستثناء التطبيقات المعتمدة على التعليم بدون معلم.

إن الشعبية الواسعة لهذه الشبكات في استعمالها لتطبيقات عملية متنوعة مستمدة من مقدرتها التطبيقية (mapping). إذا توفرت مجموعة جيدة من معطيات التدريب لمسألة ما فإنه يبدو جلياً أن شبكة MLFF بطبقة مخفية واحدة، أو على الأكثر بطبقتين مخفيتين، تستطيع التعلم لتنفيذ المهام المرغوب بها بفعالية عالية.

سنعرض فيما يلي حلول الشبكات العصبونية الصنعية التمثيلية لأربع من أشهر فئات التطبيق وهي: التصنيف والتشخيص(calassification and diagnosis)، والتحكم والأمثلية (control and optimization)، والتنبؤ والتكهن (prediction and forecasting)، وتعرُّف الأشكال (Pattern recognition)، وقد عرضنـــا مثالاً عـــن تطبـــيق ضـــغط المعطيات (Data compression) في الفصل السابق.

2.7 تطبيقات التصنيف والتشخيص

Classification and Diagnosis Applications

إن قضايا التصنيف والتشخيص واسعة الانتشار، وتحدث في حقول مسائل عديدة، مثل حالة نظام إلكترونـــي أو ميكانيكي ينفذ مسائل عملية. سيتعرض هذا النظام إلى خطأ أو أخطاء، عندها يلزم كشف الخطأ (Identification) وإصلاحه، وقبل ذلك كله نأمل بالتنبؤ بوقوع الخطأ قبل حدوثه (Prediction) لتجنب الأضرار والكوارث الناجمة عن وقوعه، يجب فحص نوعية وجودة المنتجات المصنعة لتصنف مقبولة أو مرفوضة. مثلاً في معامل الزجاج والمصابيح الكهربائية تُفحَص القطعة المصنعة عند نهاية خط التصنيع قبل التغليف مباشرة لمعرفة صلاحيتها. وهذا الاختبار ضروري في جميع المواد المصنعة آلياً. وهكذا فإن التطبيقات في مجال التشخيص لانهائية.

سنصف هنا مثالين للتصنيف والتشخيص أحدهما يستعمل في المجال الطبـــي للكشف عن سرطان المثانة، والثانـــي في الصناعة للكشف عن الخطأ وماهيته في أنظمة المقاسم الإلكترونية الهاتفية.

1.2.7 تصنيف الخلايا لتشخيص السرطان

تعتبر البنية وخواص أخرى للخلايا المراقبة من عينات بول مريض مؤشراً دقيقاً لسرطان المثانة. إن خطة تصنيف ثنائية الفئة بسيطة "سليمة" و"غير سليمة" كافية لمهمة تشخيص الخلية، وذلك عندما تختار صفات التمييز (بين الخلايا السليمة والمصابة) اختياراً جيداً. وقد قُدِّمت عدة اقتراحات لحل مشكلة التصنيف مثل: استعمال مصنفات شجرة التطبيق الانتقائية (selective mapping tree classifiers)، واستعمال الشبكات العصبونية الصنعية المتعددة الطبقات الأمامية التغذية، وتقنيات أخرى (Moallemi عام 1991 [127]).

لقد أعطت الشبكات العصبونية المتعددة الطبقات مستويات دقة تصنيف مقبولة للتشخيص السريري. مثلاً دقة مصنفات الشجرة كانت من رتبة 23% مقارنة مع الطرائق التـــي لا تستخدم الشبكات العصبونية وبنفس مستوى الأداء. أما الدقة المنجزة باستعمال MLFF ثنائية القطبية فكانت من رتبة 96% وهذه دقة كافية. أُنجز تشخيص 43 صورة مجهرية تحتوي على 597 غرض (أو خلية) لاستعمالها كمجموعة تدريب واختبار للشبكة. جرى أولاً فحص الصور بالرؤية وتصنيف الخلايا من قبل الخبراء. وكانت النتائج كما يلي: 77 خلية سليمة و520 خلية مصابة. واختير حوالي نصف الأغراض عشوائياً للتدريب وتحقيق الاستقرار لعملية الفحص. واختيرت عدة حقائق كواصفات للخلية:

1. مساحة الخلية التـــي حددت بين 100 - 400 عنصر صورة (pixel) للخلايا السليمة، وستكون هذه المساحة أوسع للخلايا السرطانية.
2. دائرية الخلية، قياس كيفية تقريب الخلية لدائرة كنسبة مساحة الخلية إلى مساحة مربع يتضمن الخلية ($\pi/4$ للدائرة الصحيحة).
3. مساحة نواة الخلية.
4. دائرية النواة.
5. نسبة مساحة الخلية الكلية إلى مساحة النواة.

استعملت هذه الواصفات كدخل للشبكة العصبونية التـــي كان لها خمس عقد دخل وطبقتين مخفيتين كل منها بعشر عقد وعقدة خرج وحيدة. فسِّرت الخلية على أنها مصابة إذا كانت قيمة خرج الشبكة أكبر أو تساوي النصف، وليست سليمة فيما سوى ذلك. اعتبرت

كل صورة مجهرية مؤلفة من 256 × 240 عنصر صورة بـــ 256 مستوى رمادي.

جرت المعالجة المسبقة للصور كما يلي: أولاً، جزئت صور المستويات الرمادية إلى 16 مقطعاً (segment) كل منها بـــ 32 × 120 عنصر صورة لأخذ خلفيات الإضاءة المختلفة وشروط الظل بالحسبان. أعطيت المقاطع بعدئذ أرقاماً ثنائية بعد تعيين مستويات العتبة من حساب الجدول التكراري وتحليل المقاطع. اختير مستوى شدة العتبة لفصل الأشياء الخلوية عن الخلفية والأشياء غير الخلوية وهذا يسمح بتقطيع الشيء (أي تشكيل المقاطع). أنجز التقطيع باستعمال خوارزمية تلوين الفقاعة التـــي تخصص قيمة 0/1 لنقاط الصورة المتجاورة المتجانسة (Ballard & Brow عام 1982 [75]).

حُسِبت الواصفات الخمس المعرّفة آنفاً وغذيت للشبكة كشعاع دخل نموذج. بعد تدريب الشبكة، نفذت اختبارات على مجموعة المعطيات، وكانت الدقة 93.4% في كشف الخلايا السليمة المقبولة و97% في كشف الخلايا المصابة، وهذا يعطي معدل مقدار خطأ 3.5% وهو مستوى مقبول للاستعمال السريري. وكان الزمن اللازم لإجراء التشخيص مقبولاً تماماً، واستغرق زمن معالجة كل صورة 2.6 ثانية وزمن تصنيف 3 ثانية، وهذا جيد جداً إذا ما قورن مع الطرائق الأخرى التـــي تستغرق 32 ثانية لكل صورة.

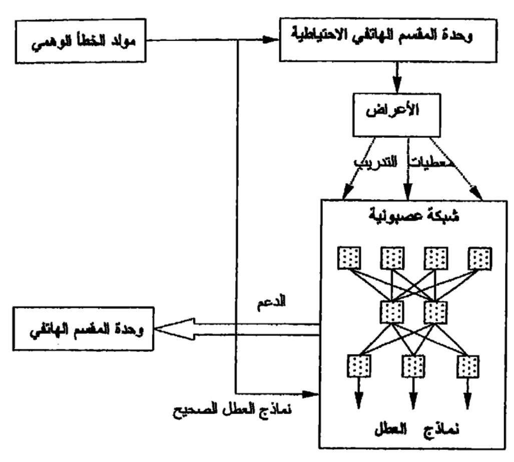
2.2.7 كشف ماهية الخطأ في أنظمة المقاسم الهاتفية

صممت أنظمة المقاسم الهاتفية الإلكترونية ESS؛ (Electronic telephone Switching Systems) الحديثة القائمة على الحاسوب لتكون عالية الوثوقية. نادراً ما تحدث الأخطاء في مثل هذه الأنظمة، ومن ثَم فإن مهندسي الصيانة لديهم حظ قليل في كسب خبرة التحري عن أخطاء التجهيزات وإصلاحها. عندما يحدث خطأ ما، تنفذ عملية التحري عن الخطأ وإصلاحه آلياً.

تجري عملية إبعاد الوحدة المتعطلة بواسطة وحدات الجاهزية الخالية من الأخطاء التـــي يستعاض بها عن الوحدات المعطلة لتكون الوحدة المعطلة معزولة تماماً. وحتـــى يمكن تحقيق تشخيص أصعب وذلك بكشف الأخطاء المتقطعة، مثل الأخطاء التـــي تحدث في أوقات متقطعة لا يمكن التنبؤ بها، أو يمكن أن تختفي قبل إتمام عملية الكشف عنها، ومن ثم تحديد طبيعة الخطأ المتقطع ستستهلك زمناً كبيراً وستكون مكلفة مالياً.

وفي محاولة لمساعدة مهندسي صيانة هذه المقاسم الإلكترونية في كشف بعض الأخطاء، دربت الشبكات العصبونية المتعددة الطبقات الأمامية التغذية لكشف مكان وطبيعة أخطاء متنوعة باستعمال معطيات متولدة بواسطة مولد خطأ وهمي (Sone عام 1993[128]).

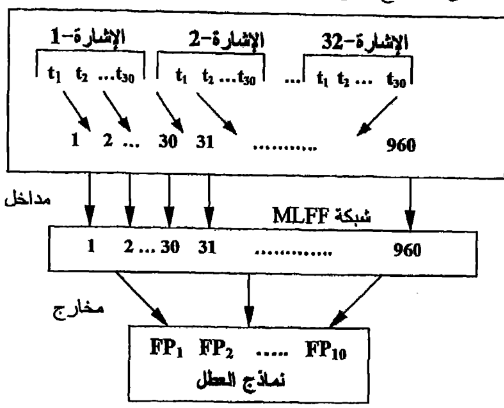
يحاكي مولد الخطأ هذا عنصر المقسم الهاتفي المتعطل بصندوق وظيفة العنصر ضمن المقسم الهاتفي الخالي من الأخطاء. بعدئذ تستعمل أعراض الخطأ الوهمي في المقسم الهاتفي التـــي أصبحت معروفة في تشخيص النظام من خلال 32 نقطة مختلفة لمراقبة الإشارات. توفر الإشارات مع الخطأ الحقيقي معطيات لتدريب الشبكة العصبونية، كما هو موضح في الشكل (1.7).



الشكل 1.7: توليد معطيات التدريب لتحديد طبيعة العطل في المقسم الهاتفي الإلكتروني

لكشف عطل فعال يجب أن تظهر كل الإشارات في مدة زمنية ما. ولتنفيذ هذه العملية، تُستخدم تقنية التحليل والتنبؤ بسلسلة الزمن العامة. يؤخذ دخل الشبكة العصبونية المشخصة للعطل من نافذة التأخير الزمنـــي المؤلفة من عدة نقاط زمنية متساوية المدد ومتتالية من

السلسلة وتستعمل كنموذج دخل مفرد.



الشكل 2.7: نماذج إشارة الدخل المأخوذة من ازياح نوافذ السلسلة

يؤخذ نموذج الدخل التالي من نفس الإشارة مع نافذة مزاحة (متأخرة) بمقدار شعاع نموذج واحد، وتكون العينة التالية متأخرة بمقدار شعاعي نموذج وهكذا حتـــى نصل إلى نهاية السلسلة الزمنية العامة. يبين الشكل (2.7) عينات التأخير الزمنـــي.

تتألف مجموعة التدريب التامة من سلسلة نماذج النافذة (الأعراض)، ونموذج الهدف الحقيقي وهو الخطأ المعروف من أجل نفس الأعراض. لاحظ أنه عندما تكون معطيات سلسلة الزمن متقطعة في الزمن الحقيقي، يجب أن تستعمل وحدات التأخير الزمنـــي المشابهة للوحدات المستعملة في تطبيقات شبكة أدلين لإنقاص الضجيج المتكيف على دارات الاتصالات.

في هذا التطبيق سجِّلت معطيات السلسلة الزمنية سلفاً، وبذلك تأخذ نقاط الزمن المتقطع مباشرة من قاعدة المعطيات المتضمنة معطيات نموذج الدخل. في التطبيق الحالي، تتألف نافذة التأخير الزمنـــي من 30 عينة بأبعاد زمنية متساوية تستعمل لكل إشارة دخل. وكذلك هناك 960 = 32 × 30 قيمة إشارة تكوّن شعاع نموذج الدخل بالكامل.

ثبت عملياً أنه في حالة عدة مسائل للشبكة العصبونية، **يكون تنفيذ عدة شبكات عصبونية صغيرة لكل عمل جزئي صغير من مسألة كبيرة فعالاً أكثر من شبكة عصبونية واحدة كبيرة لحل المسألة ككل.** ويكون زمن التدريب أقصر عندما تجزأ المسألة وتعطى إلى عدة شبكات أصغر. هذه الإستراتيجية في حل القضايا الكبيرة معروفة بمبدأ "فرق تسد" الذي يُستخدم لتحقيق مسألة تحديد طبيعة الخطأ في المقاسم الإلكترونية الحديثة.

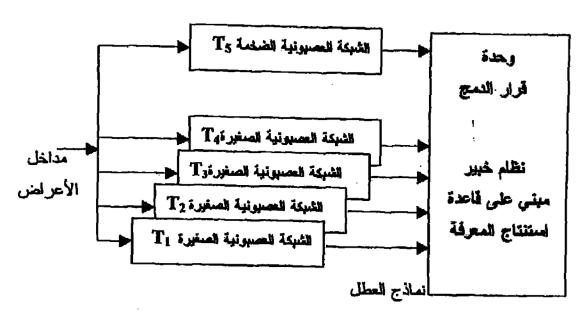
رُكِّب النظام النهائي من خمس شبكات متعددة الطبقات أمامية التغذية موزعة للتشخيص ونظام خبير تقليدي مبنـــي على قاعدة استنتاج المعرفة مدموج مع مخارج الشبكة العصبونية. أربع من هذه الشبكات العصبونية (T_1, T_2, T_3, T_4) خفضت في الحجم حتـــى 320 عقدة دخل، والشبكة الخامسة (T_5) لها 960 عقدة دخل (للأعراض). يوضح الشكل (3.7)النظام الهجين بالكامل.

الشبكات العصبونية الأربع الموزعة (T_1, T_2, T_3, T_4) لكل منها 320 عقدة دخل و600 عقدة مخفية و10 عقد خرج مع 196000 وصلة. ولشبكة الإظهار الضخمة (T_5) 960 عقدة دخل و510 عقدة مخفية و10 عقد خرج مع 495000 وصلة. لقد تحسن أداء النظام بسبب تقسيم الشبكة إلى أربع شبكات جزئية كما هو موضح في الشكل (3.7) تحسيناً واضحاً مقارنة مع استخدام شبكة واحدة ضخمة، مثلاً سجل النظام المحقق حوالي 89% تحديد ماهية خطأ بوجه صحيح على معطيات ليست تدريبية باستعمال أربع شبكات صغيرة مقارنة مع تسجيل 64% تحديد ماهية باستعمال شبكة ضخمة مفردة؛ أي بتحسين أداء بمقدار 25%.

3.7 تطبيقات التحكم والاستمثال

Control and Optimization Applications

تعتبر قضايا التحكم والاستمثال من التطبيقات التي هي أكثر صعوبة للتنفيذ باستخدام الشبكات العصبونية الصنعية. عادة تكون التطبيقات (mapping) التـــي يجب أن تُعلَّم معقدة جداً في الطبيعة، وقد تكون الشروط المقيدة التـــي يجب تحقيقها متعارضة، وحتـــى استعمال الشبكات المتعددة الطبقات الأمامية التغذية لتنفيذ بعض المسائل لاقى نجاحاً معتدلاً.



الشكل 3.7:

سنصف في هذه الفقرة ثلاثة تطبيقات هامة لشبكات MLFF في حل مسائل التحكم والاستمثال. الأول ما يزال قيد البحث والتطوير، وهو يتعلق بالتحكم بالعربات ذاتية القيادة (بدون قائد). والثانـي يتعلق بإنتاج الفولاذ، الذي صرف عليه جهد كبير لإنقاص الكلفة المالية. والتطبيق الأخير يشتمل على عدد من التطبيقات الخاصة كالتحكم الأمثل بالمنتجات الاستهلاكية.

1.3.7 العربات الأرضية ذاتية القيادة

حولت جامعة Carnegie-Mellon شاحنة تجارية إلى عربة مخبرية (Navlab I) عام 1986 لتكون كقاعدة اختبار في تجارب القيادة الذاتيـة. وقـد أسّس (Thorpe عـام 1991[76] و Kanadeعام 1994[77]). واحداً من أنظمة التحكم، المسمى عربة أرضية ذاتية القيادة بالشبكات العصبونية ALVINN، (Autonomous Land Vehicle In a Neural Networks) على شبكة عصبونية، وجُهِّزت الشاحنة بكاميرات فيديو عديدة، وكاشف مجال بالمسح الليزري، ونظام تمركز كامل، ونظام ملاحة داخلي، وحساسات شمسية، إضافة إلى ذلك فقد كانت حمولتها التـي أثقلت ظهرها عدة حواسيب حديثة. توبع الطريق الأتوماتيكي للعربة ALVINN بجهاز تحكم يستعمل شبكة عصبونية صنعية متعددة الطبقات أمامية التغذية مدربة بالانتشار الخلفي، مؤلفة من ثلاث طبقات متصلة بالكامل مع مداخل بصرية (رؤية) ملونة

من كاميرا فيديو. وقد خُفِّض تحليل صورة الدخل إلى شبكية دخل بـــ 30 × 32 عنصر صورة (أو 48 × 48 عنصر صورة).

تألفت الطبقة المخفية من 9 وحدات وطبقة الخرج من 45 وحدة، وقد جرى تعليم الشبكة بمجموعات مختلفة من الأوزان لتستطيع الشاحنة السير على طرق مختلفة. يوضح الشكل (4.7) البنية الأساسية للشبكة.

خلال العملية، عولجت الصورة سلفاً لتقوية تباين الطريق ومن ثم غذيت النتائج مباشرة لدخل الشبكة. تحسب الشبكة زاوية التوجيه مباشرة دون التفكير بمحل الطريق. وتوفر عقد الخرج تعليمات توجيه العربة بدرجات متغيرة من الدوران الزاوي؛ (45 مكان زاوي مختلف) أقصى اليسار، وتغير درجات اليسار، وإلى الأمام، وتغير درجات اليمين، وأقصى اليمين.

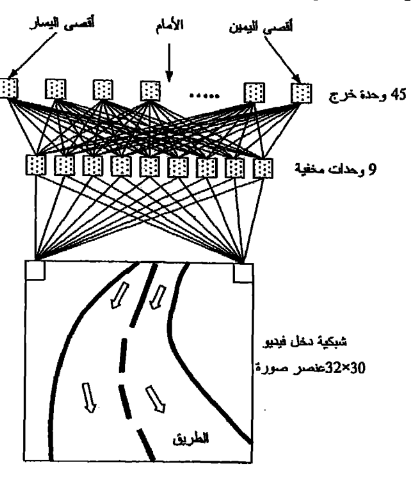
وجرى تحديث الخرج خلال العملية 15 مرة في كل ثانية، وذلك لتوفير تحكم فعال في الزمن الحقيقي للعربة وهي تسير بسرعة حتـــى 55 متر/ساعة أو أكثر.

لتدريب الشبكة استعمل إجراء وحيد "on-the-fly". فقد أدخلت الصور وعولجت في الوقت الذي كان فيه شخص يقود العربة في الطريق. تألفت مجموعة التدريب من صور فوتوغرافية لصورة الطريق العام (شعاع الدخل)، والمكان الزاوي لميكانيكية التوجيه (شعاع خرج).

أعُيد فعلياً استخدام كل صورة في أماكن عديدة خلال التدريب. وحوِّلت الصورة خلال عمليات العبور الجانبـــي لتوفير أمكنة عديدة مزاحة لمحاكاة التطابقات الخاطئة بين الطريق وزاوية التوجيه، وعيِّنت أيضاً الأوامر المناسبة لتصحيح الخطأ، وزودت كل الصور وزوايا التوجيه المعدلة كجزء من مجموعة التدريب. يعطي هذا النوع من التدريب مشهداً للطريق منسجماً كثيراً مع فنون القيادة البشرية. استُعمل أيضاً نظام الصور ثلاثية البعد من كاشفات المدى الليزرية لكشف الأشياء على طول الطريق (أشجار، سيارات، حجارة، صناديق بريد، ...الخ).

لقد صادف مشروع ALVINN نجاحاً كبيراً حيث انتقلت العربة بسرعة 55م/سا لمسافة 90 ميلاً أو أكثر. أيضاً اختبرت العربة بنجاح على أنواع طرق مختلفة(خط مفرد ومضاعف، وطريق ترابـــي ومعبد) وفي شروط مناخية مختلفة.

ليست الشاحنة ALVINN هي نظام التحكم الناجح الوحيد المستعمل للشبكات العصبونية الصنعية في قيادة العربات ذاتياً. فقد صممت وكالة مشاريع البحث



الشكل 4.7: وحدة تحكم الشبكة المتعددة الطبقات بـ NAVLAB

المتقدم ARPA (Advanced Research Project Agency) عربات ذاتية القيادة ALV أيضاً، بالإضافة إلى المنظمات الأوربية واليابانية.

إن الجهود المبذولة على تطبيقات المركبات بدون قائد، عسكرياً وتجارياً، ضخمة بالفعل وتستحق انتباهاً خاصاً ويقظة وتشمير عن ساعد الجد والعمل العربي.

2.3.7 عنصر تحكم ذكي في صناعة الفولاذ

توضع النفايات المعدنية في عملية صناعة الفولاذ في الفرن، ثم يُستعمل مسخن بالقوس الكهربائي لصهر هذه النفايات. كلفة العملية في مثل هذه الأفران كبيرة جداً، وذلك بسبب الكمية الكبيرة من الطاقة الكهربائية المستهلكة خلال عملية الصهر، والكلفة عند مستويات

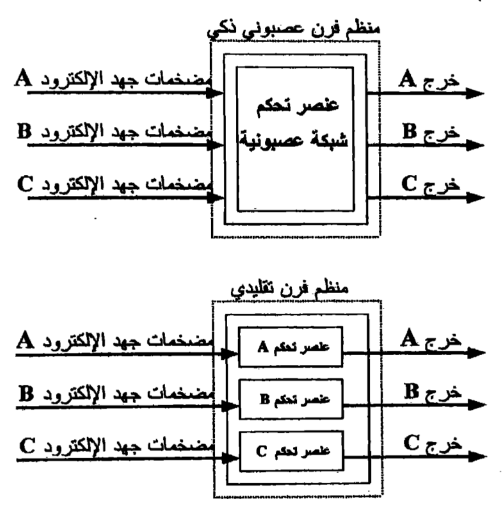
معينة أيضاً عالية بسبب الوسط الخارجي المحيط بالعملية، والكلفة العالية للمنشأة الصناعية، لذا فإن التحكم الفعال في العملية ككل سيكون حرجاً. هذا وإن تغير فعالية العملية لبضعة أجزاء من المئة ينتج عنه فروق كبيرة في التكلفة المالية للعملية.

في عملية تصنيع الفولاذ، عندما تنطلق دفعة جديدة من الحرارة، تقريباً 100 طن من النفايات المعدنية توضع في حجرة الفرن التـــي يبلغ قطرها 15-30 قدماً. بعدئذ تنخفض ثلاثة إلكترودات غرافيتية ضخمة باتجاه كومة النفايات بواسطة محركات سيرفو. قطر الإلكترودات الأسطوانية الشكل من 12-24 بوصة وطولها 20 قدماً وكل منها يمكن أن يزن عدة أطنان من الكيلوغرامات. إن الحجم الهائل لهذه التجهيزات واستهلاك عدة ميغاوات (مليون وات) من الطاقة بهذه الوحدات ينتج عنه عرض مؤثر جداً للشرارات وكبت الانفجار عند أول إشعال للإلكترودات. ففي عملية فعالة من وجهة نظر الكلفة المالية، يجب أن يكون مكان الإلكترودات الثلاثة منسقاً ومتحكماً فيه بدقة خلال عملية الصهر.

إن أنظمة التحكم التقليدية ليست مثالية لأنها تستخدم عناصر تحكم مستقلة لكل إلكترود، وكذلك تستعمل بمجموعة ممانعات مثبتة تشير لمكان الإلكترودات. ففي عملية أكثر فعالية، يجب أن يكون التحكم ديناميكياً وينظم بدقة تيار الأطوار الثلاثة في الإلكترودات خلال عملية الصهر الطويلة. لقد اخترع Bill Staib عنصر التحكم الذكي بفرن القوس الكهربائي المستعمل للشبكات العصبونية الصنعية، عام 1993[78] ليتحكم بمكان الإلكترودات في الفرن كما هو موصوف آنفاً. وقد جهزت هذه الأنظمة في عدد من منشآت تصنيع الفولاذ، وأظهرت فعالية في تخفيض الكلفة المالية بالمقارنة مع أنظمة التحكم التقليدية، مع تحسينات تصل إلى 5 - 8% في إنقاص استهلاك الطاقة الكهربائية و20% في زيادة عمر هذه المنشآت الصناعية. تقدر الكلفة الكلية الموفرة نتيجة استخدام هذه التقنيات الحديثة بملايين الدولارات سنوياً.

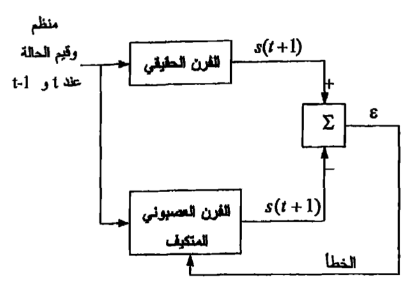
إن بنية عنصر التحكم بالفرن هي شبكة عصبونية متعددة الطبقات أمامية التغذية مدربة بخوارزمية الانتشار الخلفي. يشمل دخل الشبكة حالة الفرن (تيارات، جهود، أصوات الفرن من ميكروفون) ومخارج الشبكة هي إشارات تحكم في الأطوار الثلاثة كما هو موضح في

الشكل (5.7).



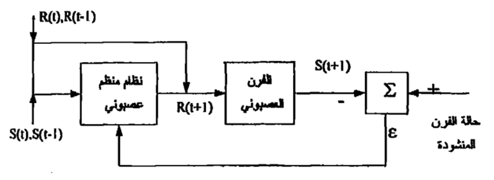
الشكل 5.7: نظام التحكم في فرن صهر الفولاذ

تتألف عملية تدريب عنصر تحكم الشبكة العصبونية من مرحلتين: تعلم الشبكة MLFF في المرحلة الأولى كيف تحاكي سلوك الفرن العملي. يقدم للشبكة والفرن نفس منظم الدخل R(t) وإشارات الحالة S(t) عند اللحظات t و t−1. تقارن بعدئذ إشارات تمثيل حالة خرج الفرن لتحاكي حالة الخرج S(t) المحسوبة بواسطة الشبكة في الزمن t +1. يستعمل بعدئذ الفرق بين خرج الشبكة المحاكاة وخرج الفرن الفعلي كإشارة خطأ ضرورية لتدريب الشبكة، كما هو موضح في الشكل (6.7).



الشكل 6.7: تدريب الشبكة لتحاكي الفرن

بعد أن يتم تدريب الشبكة لتحاكي نموذج عمل الفرن الحقيقي، تستعمل لتدريب شبكة MLFF ثانية لتعمل كعنصر تحكم منظم لفعالية الكلفة في الشبكة الأولى. في هذه الحالة، تقارن الحالة المنشودة للفرن مع حالة خرج الشبكة المحاكاة والفرق بين الاثنين يستعمل كخطأ لتدريب شبكة المنظم. مرحلة التدريب الثانية موضحة في الشكل (7.7).



الشكل 7.7: تدريب الشبكة لتتحكم بالفرن باستعمال الشبكة المدربة على محاكاة الفرن

إن عملية تدريب الشبكة لمحاكاة ديناميكيات نظام ما ومن ثم استعمال شبكة المحاكاة لتدريب شبكة أخرى لتعمل كعنصر تحكم في العملية، تعتبر مفهوماً قوياً جديراً بالاهتمام. يمكـــن أن تستعمـــل هذه العملية لعدد من التطبيقات الهامة كما وصفها Widrow عام 1994 [35].

تدعى هذه العملية تعريف النموذج (Model Identification) في أنظمة التحكم، والمحاكي (Emulator) هو المُعرِّف (Identifier).

3.3.7 أمثلية التحكم في المنتجات الاستهلاكية

كانت الشركات اليابانية وما تزال من شركات التصنيع الأولى التي دمجت تكنولوجيا المنطق العائم في منتجاهما، وقد بدأت بذلك منذ العام 1990. وبعدها جرت بالتتالي تحسينات أكثر تطوراً على العديد من هذه المنتجات المستعملة للشبكات العصبونية الصنعية. مثلاً، من بين المنتجات اليابانية المطروحة في الأسواق التـــي دمجت عناصر تحكم عصبونية صنعية: مكيفات الهواء، والمراوح الإلكترونية، ومسخنات الكيروسين، والمسخنات الكهربائية، والطباخات التحريضية، وأفران الأمواج الميكروية، والثلاجات، والمكانس الكهربائية، ومجففات الملابس، والغسالات الآلية، وآلات التصوير، ومعالجات النصوص، وغيرها كثير.

وقد استُعملت الشبكات العصبونية الصنعية بنموذجيها المستقل والمرتبط بعناصر تحكم المنطق العائم لتحسين الأداء، واستعملت أيضاً في تصميم توابع الانتماء العائمة الأمثلية لأنظمة التحكم العائم.

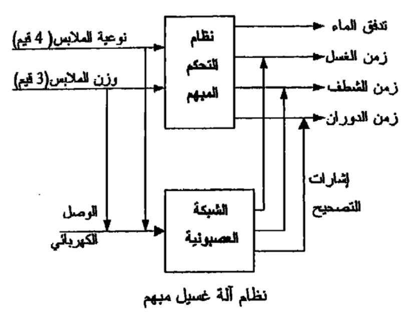
فقد طورت شركة Mitsubishi الإلكترونية اليابانية ثلاجة يُتَحكم فيها بواسطة تقنيتـــي الشبكات العصبونية الصنعية وعنصر تحكم المنطق العائم.

إن التحكم في الأدوات المنـــزلية الكهربائية المستخدم لكلتا الطريقتين السابقتين أصبح مألوفاً لدى أفراد العائلة، إذ يعتبر أحد نماذج الاستعمال اليومية الروتينية للعائلة. فقد قُسِّم كل يوم إلى 12 مرحلة زمنية، وأُحصي عدد مرات تكرار فتح باب الثلاجة خلال كل مرحلة. ومن هذه المعطيات، حُسِب معدل حركة اليوم الثامن ليجري على أساسه التحكم بالتغير اليومي لعمل الثلاجة. باستعمال هذه الحالة وحالات أخرى من المعطيات كدخل، استطاعت الشبكة معرفة متـــى يجب أن تبرد، ومتـــى يجب أن تبرد أكثر قبل الاستعمال اليومي المتكرر للثلاجة، ومتـــى ستزيل الرائحة، ومتـــى تعمل بضجيج منخفض (وخاصة في أوقات نوم أفراد عائلتها الكرام!) (Asakawa & Takagi عام 1994 [79]).

سيُنجَز التبريد بفعالية كبيرة عندما تثبِّت هذه الثلاجة درجةَ وزمن التبريد في مرحلة الفتح غير المتكرر وتقوم بالتبريد القبلي قبل المرحلة التـــي يتكرر فيها فتح الباب. وقد أدى هذا إلى إنقاص تغيرات درجة حرارة الطعام حوالي 2,2 درجة مئوية فقط. هذا وتستعمل أيضاً أداة التعليل العائم معطيات تردد فتح الباب، وحساساً لكشف كمية الجليد لتمارس تحكماً كاملاً في الثلاجة.

بعد زحف تكنولوجيا عناصر التحكم العائمة إلى المنتجات الاستهلاكية وتربعها على عرش الأولويات لدى أفراد العائلة، بحثت الشركات اليابانية عن تحسينات أخرى، لزيادة الخدمات المقدمة للعائلة، وذلك بدمج المعطيات من عدة حساسات لجعل عملية التحكم أكثر نعومة وحساسية. بالطبع هذا سيزيد تعقيد النظام وتدريبه بالتناسب مع زيادة في أبعاد فراغ الدخل. في محاولة التصدي لهذه المصاعب كانت الشبكات العصبونية الأداة الأقوى والأجدر في معالجة مجموعة ضخمة من معطيات حساسات متنوعة.

في أول نظام، استعملت الشبكات العصبونية لتصحيح خرج عنصر التحكم العائم الذي صمم أساساً لمجموعة مختلفة من المداخل. وهذا أبْعَدَ الحاجة إلى تصميم نظام كامل، وأعطانا توفيراً كبيراً في الزمن والنفقات. هناك حالة أخرى ــ ولنفس الغرض ــ استُعملت لآلة غسيل شركة Hitachi اليابانية. النظام الأول دمج مع النظام العائم كما هو موضح في الشكل (8.7).



الشكل8.7 : شبكة عصبونية تربط خرج عنصر التحكم العائم

فيما بعد ظهر نموذج محسن مع حساسات إضافية لقياس شفافية الماء. لتأمين وصلات المداخل الإضافية بدون إعادة عملية التصميم بالكامـــل، أضيفت شبكـــة عصبونية مُصَحِّحَة (الصندوق الأسفل في الشكل (8.7))، أعطيت المداخل الإضافية إلى الشبكة فقط لكي تعزز التحكم في الغسالة وتصحح خرج أداة التعليل العائم. هناك منتجات يابانية أخرى مطروحة

في الأسواق تضم خواص الشبكات العصبونية الصنعية وأنظمة التعليل العائم.

مثال أخير: طورت شركة Toshiba فــرناً يعمــل علــى الأمــواج الميكروية (بتردد $10^{+9} \times 2.45$ هرتز تقريباً) يستطيع تقدير عدد المواد المطبوخة ودرجة حرارة الفرن. تقرر الشبكة عندئذ زمن الطبخ الأمثل باستعمال التعليل العائم. يعتمد زمن الطبخ على عدد الأنواع في الطباخ، ودرجة حرارتها الأصلية، وهذا يجري تحسسه وتقديره بشبكة عصبونية تنظم تدفق غاز الطباخ. يمرر خرج هذه الشبكة إلى عنصر التعليل العائم الذي سيحدد زمن الطبخ والطاقة (الحرارة) اللازمة.

4.7 تطبيقات التنبؤ والتكهن

Forecasting and Prediction Applications

التنبؤ مهمة كل منظمة تريد أن تَعلم لتفعل. مثلاً، تريد شركات المنتجات الاستهلاكية معرفة النمو في مبيعات إنتاج جديد لطرحه في الأسواق، ويحتاج علماء الطقس إلى التنبؤ عن الطقس، وتريد المصارف التنبؤ عن استحقاقات الأرصدة للشركات كأساس للقروض الممنوحة، وتريد مجموعة إدارة المطارات معرفة نمو وصول المسافرين إلى المطارات المشغولة، وتريد شركات الطاقة الكهربائية معرفة تزايد طلبات المشتركين من الكهرباء خلال العقود القادمة لتتمكن من توفير محطات التوليد التــي تلبــي هذه الطلبات، وهكذا.

لقد أظهرت الشبكات العصبونية الصنعية نجاحاً فائقاً كأدوات تنبؤ بطرق مختلفة؛ مثل التنبؤ بأن حادثاً ما سيقع أو لن يقع، والتنبؤ بالزمن الذي سيقع عنده هذا الحادث، والتنبؤ بمستوى تأثير حدوث هذه الحادث. يجب أن تدرب الشبكة العصبونية الصنعية جيداً لكي تستطيع التنبؤ بمستويات دقة مقبولة، وأن تطبق على عدد ضخم من الأمثلة المستقاة من النماذج الماضية مع قيم التأثير المستقبلية المعروفة. عادة تأتــي مجموعة التدريب من معطيات تاريخية مُجمَّعة عبر مدة معطاة، ويجب على الشبكة العصبونية أن تتعلم التعميم والاستقراء من نماذج جديدة للتنبؤ بالتأثيرات المستقبلية. سنصف في هذه الفقرة نوعين من تطبيقات التنبؤ باستعمال شبكات MLFF هما دمج معطيات حساس والتنبؤ بالخطأ، والتنبؤ

باستحقاقات الاعتمادات في التطبيقات المالية.

1.4.7 دمج معطيات حساس والتنبؤ بالخطأ

Data Fusion and fault prediction Sensor

إن عملية دمج معطيات حساس هي عملية ضم معطيات حساس متعدد المنابع لأهداف الكشف، والارتباط، وتعريف النموذج (Identification)، والتنبؤ، وتقدير الوضع. وقد أنجزت عدة تقنيات دمج للمعطيات وخاصة في العشرين سنة الماضية، وتطور عدد من التطبيقات بخطى سريعة.

إن التقنيات التـــي تستعملها الطرق الحالية تشمل اختبار الفرضيات الإحصائية، وتقنيات ضم بايز (Bayes)، والطرق الاحتمالية الأخرى (التعليل الواضح لـــ Dempster-Shafer)، والمنطق المبهم، ومقدار الكفاءة، والأنظمة الخبيرة التقليدية، والشبكات العصبونية.

تشمل التطبيقات النموذجية الكثيرة والمتعددة للشبكات العصبونية الصنعية في دمج معطيات الحساسات: أنظمة القيادة، والاتصالات، والاستطلاع (التجسس)، والذكاء، وكشف الأخطاء الأولية للأنظمة الإلكتروديناميكية المعقدة، والربوت، وعمليات المراقبة، والتحكم،...الخ.

وقد أثبتت دراسات مقارنة أن الشبكات العصبونية ذات إنجاز أفضل من الطرق التقليدية في مجال عدة تطبيقات دمج. وأحد هذه المجالات الواعدة كثيراً في مجال تطبيق الشبكات العصبونية الصنعية هو التنبؤ بالخطأ المهدد بالوقوع (التنبؤ بخطأ أولي) الذي ينتج عنه عادة الكثير من التصليح المكلف، أو الكوارث وفقدان الحياة.

مثلاً علبة السرعة الأساسية في الطائرات المروحية (Hilocopter) ستتعطل ميكانيكياً بمعدل 300000 ساعة عمل. ينتج عن مثل هذا العطل أو الخطأ في علب السرعة، والأبراج، وأدوات حرجة أخرى للقيادة، كوارث مالية وبشرية لا تحمد عقباها على الأغلب. ومن ثَم فإن التنبؤ في مثل هذه الأخطاء والتكهن بها بدقة قبل وقوعها يُمَكِّن العربة (أرضية أو جوية) من التوقف والهبوط بأمان لإصلاحها وتجنب إصلاح رئيسي أكبر وضرر خطير لطاقم القيادة

فيما إذا وقع الخطأ.

أظهرت الفحوص المخبرية للطائرات المدنية (Boeing) أن كشف فرقعات (أصوات غريبة) التعب للقطع الميكانيكية في المراحل الأولى ممكن عندما تكون 10% فقط من مساحة الجزء المتعب قد تصدعت. أي إن الكشف في الزمن الحقيقي وكشف الخطأ قبل وقوعه ممكن من خلال تحليل إشارة الاهتزاز. من ناحية أخرى، إن أنظمة المراقبة التقليدية العديدة غير مقبولة في هذا المجال لأن أجهزتها ذاتها قابلة للعطل (وتحتاج إلى مراقب لها) أكثر من الأنظمة التـــي تراقبها. فمثلاً، معدل تعطل الحاسوب هو 3000 ساعة وهذا المعدل أكثر بـــ 100 مرة من معدل تعطل علبة السرعة ذاتها المراقبة بذاك الحاسوب. لذا نحتاج إلى أنظمة مراقبة دقيقة وموثوقة، تستطيع القيام بالتشخيص في الزمن الحقيقي وعلى ظهر المركبة من خلال دمج معطيات العديد من الحساسات: كنظام القدرة، ونظام القيادة، والبنية، والنظام الدوار، وأنظمة الزيت.

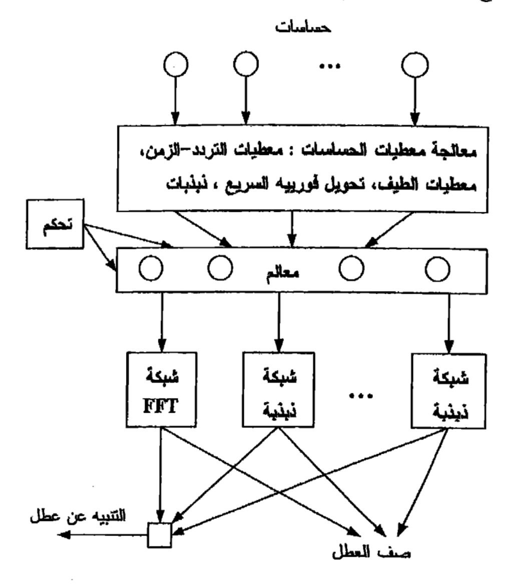
يظهر بعدئذ أنه إذا أمكن إنتاج شرائح (chips) الشبكات العصبونية الصنعية المُشخِّصة المنخفضة الكلفة بوفرة، فإن هذه الشرائح ستجد لها تطبيقات هائلة. ويمكن أن تستعمل بكلفة معقولة للمحافظة على أية عربة أرضية أو جوية أو بحرية تتطلب العمل المستمر المكلف.

لقد انصبت جهود عدد من الوكالات العسكرية والمدنية على هذا النوع من التشخيص، وقد بدأت المطالب باستخدام تكنولوجيا مـــراقبة الصحـــة والاستعمال HUM (Health and Usage Monitoring)، وخاصة أن لدى مراكز البحوث البرامج الأولية لتنفيذ مثل هذه التكنولوجيا. مثلاً، طلبت المملكة المتحدة والنروج من جميع الطائرات المروحية التـــي تقوم بعمليات طيران عبر بحر الشمال أن تكون مزودة بتقنية مراقبة الصحة والاستعمال HUM على ظهرها بدءاً من عام 1995 (Rock وزملاؤه عام 1993[80]).

تقنية مراقبة الصحة والاستعمال:

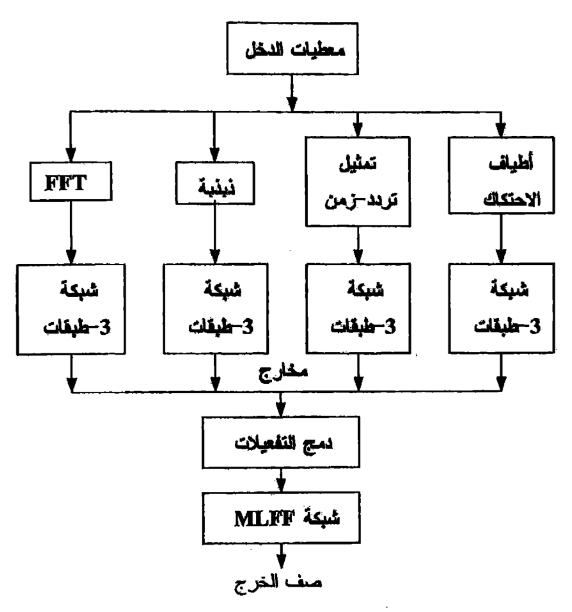
إن الهدف من برنامج مراقبة الصحة والاستعمال للطائرات المروحية هو تطوير النقل الجوي في الزمن الحقيقي؛ ويعتمد على طريقة تحليل إشارة الاهتزازات المبثوثة لتشخيص

موثوق لأعطال تعب القطع قبل حدوثها. وبذلك نضمن هبوطً آمناً اضطرارياً قبل حدوث العطل. يوضح الشكل (9.7) نظام مراقبة الصحة والاستعمال.



الشكل 9.7: مخطط نظام مراقبة الصحة والاستعمال في الزمن الحقيقي

يتألف النظام من حساسات متعددة (مقاييس تسارع، حساسات صوتية، مقاييس قوة إجهاد)، ومعالجات مسبقة للإشارة، وواحدة أو أكثر من شبكات التنبؤ العصبونية المتعددة الطبقات، وأجهزة إخراج متنوعة. يبلغ عدد ملامح (خواص) الدخل المعالج من قبل الشبكة نموذجياً من 10 -20.



الشكل 10.7: تسلسل النظام لاستنباط معالم الداخل، دمج وتصنيف

يمكن تقسيم هذه المداخل وتوزيعها على شبكات استنباط الملامح قبل أن تدمج بواسطة شبكة المصنِّف التنبئي، أو تدمج وتشخص بواسطة الشبكة العصبونية الكبيرة المفردة. يوضح الشكل (10.7) البنية لهذا النوع من المعالجات (Brotherton & Pollard عام 1992[28]).

لقد استعملت هذه البنية في سلسلة من الاختبارات الواعدة، حيث دمجت مخارج شبكات الطبقة الأولى مع شبكة الطبقة الثانية لمعالجة عملية الدمج. يسمح هذا النوع من المعالجة التسلسلية التصاعدية بأزمنة تكامل إشارة طويلة تعطي شبكة الخرج زمناً للتعلم حيث تكون المداخل مفصولة أو تنذر بالخطأ، ويمكن أن تكون متجاهلة.

عند تدريب الشبكة، وجد أن 20 دورة فقط لعمود دوران المروحة تكفي لتوليد معطيات

كافية لتباينات الدقة بين حالات الإرسال المختلفة. رغم التجارب العديدة لأنظمة HUM المستعملة للشبكات العصبونية الصنعية لدمج معطيات الحساسات والتنبؤ بالخطأ المهدد بالوقوع فإنها ما تزال محدودة. ومن المحتمل جداً أن شرائح الشبكات العصبونية الصنعية ستصبح متطلبة لأنواع عديدة من العربات الجوية والبحرية في المستقبل.

2.4.7 التنبؤ باستحقاقات الاعتمادات للتطبيقات المالية

يعد التطبيق المقترح والمستعمل من قبل مصرف (Chase Manhatten) واحداً من أعظم التطبيقات الناجحة لنماذج التنبؤ باستعمال الشبكات العصبونية . كان النظام هجينياً: شبكة عصبونية مبنية إحصائياً تحدد استحقاقات الاعتماد للشركات العامة الباحثة عن قروض تجارية.

يقرض هذا المصرف حوالي 300 مليون دولار سنوياً لتوظيف الشركات. وحيث إن هذا المبلغ يمكن أن يكون مصدراً كافياً للربح (أو للخسارة)، فإن عملية التكهن الدقيق في قيم اعتمادات أو أرصدة الزبائن المقترضة ستكون أساسية جداً وهامة للمصرف.

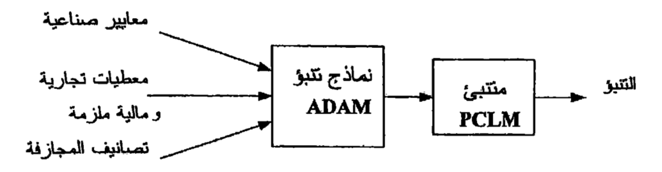
ينفذ هذا النظام المعروف باسم (Creditview) تكهنات لثلاث سنوات لتعيين مستويات المجازفة الثلاثة: **جيدة وحرجة وخاسرة**. يوفر النظام أيضاً قائمة مفصلة عن البنود ذات التوزيع الكافي للتكهن. ويفسر نظام خبير تقليدي البنود ويعطي تقارير مقارنة متنوعة لموظفي القروض ذوي المراتب العليا.

يستقبل نظام الـــ (front-end) للشبكات العصبونية المعروف باسم ADAM معطيات الدخل التاريخية من قاعدة معطيات مالية بالاشتراك مع معطيات الملتزم (obligor) الجيدة والسيئة. وتخدم هذه المعطيات كجزء من مجموعة التدريب. يستقبل ADAM أيضاً دخلاً بمعايير صناعية، ومعطيات مالية لتعميم الفئات الصناعية الخاصة.

يولد ADAM متحولات ترشيح (تزكية أو تقييم) يمكن أن تشير للحالات المالية المستقبلية للشركة. هذا يترجم كمعطيات معالم في شكل أشعة الملامح المنطقية التـــي تكون الأساس لنموذج شبكة عصبونية للشركة تحت البحث والمساءلة.

تعطي الشبكة العصبونية المستخدمة كمتنبئ ، والمعروفة بمتنبئ نموذج الشركة العام

PCLM (Public Company Model Forecaster)، معدلاً للشركة طالبة القـــرض. الشكل (11.7) يوضح النظام ككل.



الشكل 11.7: مصنف اعتماد شبكة عصبونية هجين لمصرف Chase

يقبل نظام PCLM ست سنوات من المعطيات المالية الماضية للشركة الواجب إعطاء معدل تصنيف لها. ومن ثم تستعمل التعابير الناتجة من ADAM للتنبؤ بالصحة المالية للشركة لثلاثة أعوام مستقبلية.

تعطي التنبوءات أرجحية للشركة المصنفة بمعدل جيد على باقي الشركات المصنفة بمعدل حرج أو خاسر. بالإضافة إلى ذلك، يعرف PCLM صفات القوة والضعف في البنية المالية للملزم. ويقدم إلى المستعمل تقارير موسعة مقارنة لتقديرات الخطر وتعليل نصي للتحليل.

كتب Morose عام 1993 تقريراً أكد فيه أن مصرف Chase اختبر النظام اختباراً مكثفاً وكشف النقاب عن عدد من القروض المزعجة. وضع النظام في الخدمة عام 1990 ووسع وعزز ليشمل تقييمات التعاون الخاصة أيضاً [81][82].

5.7 تطبيقات تعرف الأشكال

Pattern Recogniton Applications

يشمل تعرُّف الأشكال تطبيقات تمييز الكلام، والكتابة بلغات مختلفة، وتمييز الصور المرئية، والتصنيف، وأنواع متنوعة للإشارة، والتحليل البيانـــي مثل المخططات البيانية الكهربائية للقلب والدماغ، والتحليل التخطيطي المتنوع لعملية مراقبة الخطر.

إن تطبيقات تعرف الأشكال وتمييزها تكون عادة صعبة المراس، وهي على الأغلب

مرتبطة مع المهام المعرفية المنجزة من قبل البشر.

سنصف في هذه الفقرة ثلاثة من تطبيقات تعرُّف الأشكال وتمييز النماذج وهي: تعرُّف الكتابة اليدوية، وكشف نوبات الصرع، والتحديد الأتوماتيكي لسمات الأشخاص.

1.5.7 تعرُّف أحرف الكتابة اليدوية

إن تطبيقات تعرُّف أحرف الكتابة اليدوية آلياً لفتت الانتباه عبر العقد الماضي. وبسبب صعوبتها أصبحت علامة أداء للعديد من تصميمات الشبكات المختلفة. هذا المسعى والجهد سيكون بالتأكيد مبرراً لأن هناك تطبيقات صناعية وتجارية ضخمة تؤدي عملية التمييز الأتوماتيكي فيها إلى توفير في الكلفة.

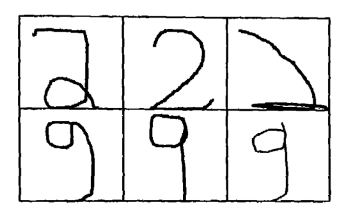
لقد درست جميع طرق التقريب المتنوعة بما فيها الطرق التقليدية لتعرُّف الأشكال والنماذج بالإضافة لحلول الشبكات العصبونية الصنعية. وتحققت درجة عالية من النجاح في المجالين الصناعي والتجاري، وتُستعمل أنظمة الشبكات العصبونية الصنعية التجارية الآن على نطاق واسع (Schwartz عام 1992 [28]). بعد هذه النجاح، من المحتمل أن تصبح أنظمة الحالة الصلبة التجارية المنخفضة الكلفة بعد وقت قريب متوفرة في الأسواق. وربما تكون الشبكات العصبونية الصنعية المتعددة الطبقات الأمامية التغذية هي الأكثر استعمالاً من بين بنـــى الشبكات العصبونية، والفروق في التصاميم ستكون على نحو أساسي في نوع معالجة الصورة وتنقية أو تهذيب بنية الشبكة.

وصْفنا فيما يلي مبنًـــي أساساً على النتائج المنجزة في مخابر (AT&T Bell) المقدمة من قبل le Cun عام 1987 [83] و1990 [84]. في تعرُّف أحرف الكتابة اليدوية، كانت الشبكية المخططة التـــي استعملت من قبل الباحثين مصفوفة بحجم 16 × 16 عنصر صورة مع 16 أو أكثر مستوى رمادي لكل عنصر صورة.

بوجه عام، يلزم مراحل متعددة من المعالجات القبلية. من الضروري أولاً، تخصيص مقطع لكل حرف، وهذا يعنـــي عزل واستنباط الحرف من النص إذا كان ذلك ضرورياً. بعد ذلك توضع للصورة قيم عتبة، وتعطى بالشكل الثنائي الرقمي المكافئ؛ أي إن نقاط الصورة ستعطى قيمة الصفر أو الواحد. أخيراً، من الضروري وضع معيار ما لإعطاء حجم قياسي

للأحرف. يؤخذ حجم الصورة المعياري عادة 16 × 16 = 256 نقطة صورة. بعدئذ، لكي تستعمل صورة الحرف المعياري المعالج كدخل فإن الشبكة سيكون لها 256 عقدة في طبقة الدخل.

لقد حققت بنية شبكة AT&T الأولى إنجازاً جيداً في تعرف الأحرف وتمييزها كتلك الموضحة في الشكل (12.7).



الشكل 12.7: أحرف نموذجية موضوعية في مقاطع للتمييز

كانت هذه الشبكة رباعية الطبقات، أول طبقتين مخفيتين تنفذان استنباط المَعْلَم، على حين تعمل الطبقتان الأخيرتان كمصنفات خانة. لقد خفض العدد الكلي لوسطاء الشبكة (الأوزان) بفرض قيود على الشبكة، أي افتراض شكل من الأوزان المشتركة: خطة تستقبل فيها العقد مداخل من أجزاء مختلفة من مستوى الدخل، لكن باستعمال نفس قيم الأوزان.

تتألف الطبقة المخفية الأولى من 12 مستوي أو مخطط (maps) مَعْلم بـــ 64 عقدة كل منها موصل جزئياً. يوافق كل مستوٍ كاشف مَعْلم مختلفاً. للعقد ضمن المخطط المعطى حقول مستقبلة محلية مختلفة، لكنها تشترك بنفس قيم الوزن.

الطبقة المخفية الثانية ذات بنية مشابهة، لها 12 مستوياً بـــ 16 عقدة كل منها متصل جزئياً. الطبقتان الأخيرتان (الثالثة والرابعة) متصلتان اتصالاً كاملاً بـــ 30 و10 عقدة على الترتيب. استعمل هذا الوصل الاختياري من قبل Fukushima في شبكة Neocognitron (سيتم شرحها فيما بعد) لتنفيذ تعرّف الأشكال المزاحة اللاتغيرية. استعمل Keeler وزملاؤه عام 1992 [85] وLee عام 1991 [86] أيضاً حقولاً مستقبلة محلية (وصلات من محلات

مكانية مختلفة) مع أوزان مشتركة في شبكات MLFF لتطبيقات تعرّف أحرف الكتابة اليدوية.

البنية المستعملة من قبل Keeler وزملائه قادرة على تنفيذ عملية تشكيل مقاطع لحظياً، وتعرّف الأشكال بمعالجة صغيرة للصورة. كان هناك عدد من قواعد المعطيات لتدريب واختبار الشبكات العصبونية الصنعية في هذه المسألة.

أحد قواعد المعطيات هذه، كان قاعدة معطيات الخدمة البريدية للولايات المتحدة الأمريكية، التـــي تتألف من 9000 خانة رمز مكتوب يدوياً (قاعدة معطيات رمز مكتوب يدوياً OAT عام 1987 [28]). تقليدياً، يتم تجزيء عشوائي لمنبع المعطيات إلى مجموعتين: مجموعة تدريب تحوي 80% من قاعدة المعطيات ومجموعة تضم 20% للاختبار. بعدئذ يكون إنجاز الشبكة المدربة مبنياً على أعداد من التصانيف الصحيحة وغير الصحيحة، آخذين بعين الاعتبار عدد مرات الرفض المسموح به عبر معطيات مجموعتـــي التدريب والاختبار.

وقد نشرت قيم إنجاز نموذجية كانت في مجال 0,14% تصانيف غير صحيحة في مجموعة التدريب و5% تصانيف غير صحيحة في مجموعة الاختبار عندما استعمل معدل خطأ 1%. لقد أنجز معدل الخطأ هذا عندما سمح بـــمعدل رفض 12% فقط.

على الرغم من تحقيق درجة عالية من النجاح في تعرّف أحرف الكتابة اليدوية (3-4% من الحروف غير المرفوضة)، ودقة تمييز عالية، فإنه يبقى هناك كتابات يدوية سيئة لا حصر لها بدون حل. أعطت تجارب حديثة في تعرّف أشكال صور ممسوحة من صفحة مكتوبة يدوياً كتابة سيئة وبدون شروط مقيـــدة دقـــة تمييـــز أحرف في المجال مـــن 61% حتـــى 81% (Yanicoglu & Sandon عام 1993 [87]). كان لأول طبقة مخفية في شبكة MLFF المستعملة في التجارب 1000 عقدة دخل موافقة لصورة عولجت من قبل بـــ 20 × 50 عنصر صورة بمقياس رمادي. الشبكة كان لها 70 عقدة طبقة مخفية و26 عقدة خرج، واحدة لكل حرف طباعة أبجدي. كانت عقد الطبقة المخفية متصلة مع الصور الجزئية لأحرف المقاطع (للأحرف الموضوعة في مقاطع) بـــ 10 × 10 حقل مستقبل دخل، وكانت متصلة اتصالاً كاملاً مع عقد الخرج.

في مجموعة واحدة من التجارب، دربت الشبكة باستعمال 28 أبجدية من 26 حرف

مكتوب يدوياً بواسطة 28 كاتباً مختلفاً. تألفت مجموعة الاختبار من سبع أبجديات مختلفة مكتوبة من قبل سبعة كتاب مختلفين. في هذه الحالة، كان معدل التمييز والتعرف 75% على مجموعة الاختبار بعد أن دربت الشبكة على تمييز كل أبجديات التدريب. في تجارب أخرى، كانت دقة التمييز أعلى أو أدنى من هذا المستوى بمعدل دقة حوالي 61%. تحقق جزء من النجاح في هذه التجارب تبعاً لخطوات معالجة الصورة السابقة التي استعملت طرقاً جديدة في تقطيع الكلمات إلى أحرف، ووضع معيار لحجم وميل الأحرف الموضوعة في مقاطع.

2.5.7 الكشف عن نوبات الصرع

الكشف المبكر لنوبة الصرع قبل حدوثها يجعل من الممكن البدء بالمعالجة مباشرة. ولكن كشف الفترة السابقة للنوبة ليس أمراً سهلاً، وذلك بسبب التنوع الكبير في المعطيات الزمنية من مريض إلى آخر، وسلوك النوبة أيضاً يعتبر مصدراً للتنوع. فإذا توفرت المعطيات المختلفة على النوبات السابقة لمريض خاص، فإن استعمال تكنولوجيا الشبكات العصبونية الصنعية ممكن لأنه يمكن أن تنشأ مجموعات معطيات التدريب.

تجمع معطيات المريض بغرس أقطاب EEG لفترات زمنية جزئية في المريض لتسجيل إشارات الدماغ التشابهية عبر فترة زمنية سابقة للنوبة. سُجِّل تخطيطياً عدد من الإشارات (100 - 200 هرتز) من محلات مختلفة من أجل التحليل الكهربائي بعلم الوظائف، وجُعلتْ رقمية وخزنت للاستعمال فيما بعد في تدريب شبكات MLFF. يستطيع علم الوظائف تحليل وطباعة المعطيات باستعمال محرر شاشة تخطيطي، وإزالة الأشياء المصنعة ومعطيات أخرى خارجة عن الموضوع. نُفِّذت معالجات سابقة أخرى على المعطيات لإنقاص حجم مجموعة المعطيات.

استُعملت نوافذ منزلقة عبر السلسلة الزمنية للمعطيات لاختيار مقاطع من معطيات النموذج (عينات من معطيات النموذج). وتُستعمل هذه المعطيات في حساب المتوسط، والتباين، والجداول التكرارية، وملخصات إحصائية أخرى. قُسِّمت بعدئذ المعطيات من القنوات المختلفة بين عدة شبكات MLFF لأداء التدريب، واستُعملت مخارج هذه الشبكات كدخل لشبكة التنبؤ النهائية، وذلك بضم مخارج الشبكات العديدة المستقلة إلى شعاع يكون

دخلاً للشبكة النهائية. هذا وقد أصبح الباحثون قادرين على تحسين نوعية الاستجابة لنظام الكشف المبكر (Hamilton & Hufbael عام 1992 [88]).

3.5.7 التعرف الآلي لملامح الأشخاص

تطبق الطرائق الأتوماتيكية لتحديد ملامح الأفراد على نحو واسع في مجالات الأمن ومنع الجرائم. تُضبط الأنظمة اليومية المتكررة للأشخاص خلال العمل ويتحكم فيها من خلال التحقق من شخصية الفرد للسماح له بالتسهيلات والوصول إلى بنك المعلومات. بُنِيَتْ تقنيات التحقق العام التقليدية على استعمال كلمات السر أو بطاقات شخصية للأفراد.

سنصف فيما يلي طريقة تمييز مبنية على دمج معطيات الصورة والكلام من عشرة أشخاص (Colombi عام 1993 [89]). بنـــي التصميم على استعمال معطيات التدريب المجموعة خلال عشرة أيام لكل شخص، حيث جُمعت معطيات النطق والصورة الوجهية. ثم سجلت جميع معطيات الكلام، كانت جملاً لفظية قوية، مع اسم المُدَّعي (الشخص المقلد). وبنيت معطيات الصورة للأشخاص العشرة من أكثر من 2000 صورة مقياس رمادي. حيث كانت كل صورة 32× 32 عنصر صورة بـــ 256 مستوى مقياس رمادي. ثم جُمعت أعداد متساوية من الصور المنشودة وغير المنشودة في قاعدة معطيات التدريب. استعملت معطيات عدة أيام للتدريب بغية تحسين إنجاز التمييز بواسطة التعويض عن التغيرات اليومية في المظهر الشخصي ونماذج الكلام. وقد أعطى تدريب يومي متعدد تعميماً أفضل ومقدرة على ملائمة التشوهات.

استعملت معطيات الصورة الجبهية لتدريب شبكة MLFF بطبقة مخفية وحيدة لها 1024 عقـــدة دخل وعشر عقد طبقة مخفية وعقدتا خرج. يوافق المخرجان الاستجابات "أنا" و"ليس أنا". دربت الشبكة العصبونية على عينات معطيات تدريب وقد أنجزت دقة 100%. وعلى الرغم من أن أنظمة الشبكات العصبونية الصنعية لتمييز الكلام كانت ناجحة، فإن معطيات الكلام المستعملة في هذه التجارب كانت معالجة باستعمال تقنيات ليست عصبونية. كانت هذه التقنيات مبنية على الطرق التقليدية المعروفة باسم الترميز التنبئي الخطي وتمثيلات النموذج العصبونـــي السمعي والبصري (Colombi 1993 [89]). وكان التحقق من هذه

الحالة مبنياً أيضاً على شكل خطة "أنا" و"ليس أنا". بعد التدريب المستقل للكلام والتحقق من الصورة، دمج المخرجان باستعمال طريقة احتمالية مثقلة للحصول على تثبت نهائي لشخصيات الأشخاص العشرة.

نقلت تجارب التثبت الوجهي باستعمال شبكتين مختلفتين MLFF لنوعين من ملامح الدخل. وهي معطيات عنصر صورة سطر معمم كما وصف آنفاً، وعوامل المركبة الأساسية (المعروفة أيضاً بتحويل Karhunen-Loeve) المستنبطة من الصورة الجبهية.

تتألف الشبكة المستعملة لعوامل المركبة الأساسية كدخل من عشر عقد دخل، و20 عقدة مخفية وعقدتـــي خرج. وجرى الحصول على إنجاز أفضل إلى حد ما باستعمال مداخل المركبة الأساسية على طريقة دخل عنصر صورة سطر. كان الاختلاف في الإنجاز أعظمياً في حالة تدريب قصير الزمن. عندما استعملنا يومَي تدريب مع تسعة إلى عشرة أيام فترة اختبار، كانت طريقة دخل المركبة الأساسية ذات دقة مضاعفة (دقة تمييز 47.2% أصبحت 90,6%). وفي حالة فترة تدريب أطول، كانت طريقة دخل المركبة الأساسية أفضل إلى حد ما. مثلاً، عندما خصصت للتدريب الأيام السابع والثامن والتاسع مع الثالث والثانـــي والأول للاختبار على الترتيب، فإن إنجاز دخل عامل المركبة الأساسية كان أفضل بقرابة 10% (86% أصبحت 98,3%، و89,5%أصبحت 99%، و90% أصبحت 100%).

في مجموعة التجارب الثانية، استعملت عشر عقد خرج في شبكات MLFF، وخصصت عقدة واحدة لتحديد سمات كل شخص. للشبكة 1024 دخل معطيات عنصر صورة سطر معمم، وتستعمل طبقة مخفية بـــ 20 عقدة، على حين تتضمن الشبكة بطريقة المركبة الأساسية كدخل طبقة مخفية بـــ 40 عقدة. دربت كلتا الشبكتين حتـــى إنجاز دقة 100% على معطيات التدريب، فكانت دقة معطيات الاختبار فقيرة إلى حد ما عندما خصص يوم واحد حتـــى تسعة أيام للتدريب، وخصص اليوم العاشر للاختبار. أنجزت شبكة دخل عنصر صورة سطر معمم دقة 58,9% بالمقارنة مع دقة 87,8 لشبكة دخل المركبة الأساسية.

دليل المصطلحات العلمية

إنكليزي-عربي

-A-

Activation	تفعيل
Adaptive	متكيف
Algorithm	خوارزمية
Ambiguous	غامضة
Analog-digital	تماثلي رقمي
Antiderivative	عكس المشتق
Argument	محاكمة ـــ مُحدِّد
Artificial	صنعي
Association	اقتران- ترافق
Associative recall	استدعاء مترافق
Associative property	خاصة تجميعية (اقترانية)
Attractor	جاذب
Autoassociative	ترافق ذاتي
Autocorrelation	ارتباط ذاتي
Autonomous	ذاتي القيادة
Axon	محور

-B-

Backpropagation	انتشار خلفي (عكسي)
Basin of attraction	حوض التجاذب
Batch learning	تعليم دُفعّي
Bias	انحياز
Bidirectional	ثنائي الاتجاه
Binomial	ثنائي الحد
Biology	بيولوجيا ـــ علم الحياة
Biomolecular	جزئ حيوي
Bipolar	ثنائي القطبية

Bit	بت ـــ خانة ثنائية
Brain	دماغ

-C-

Cardinality	رئيسي / أصلي
Cascade	متتابع
Category	فئة
Cell	خلية
Chain rule	قاعدة السلسلة
Chaos	فوضوي
Characteristic	مميِّز
Chip	شريحة ـــ رقائق
Chunking	تكتل
Clamped mode	نمط الإلزام
Class	صف
Classification	تصنيف
Cluster	قطاع ـــ عُنْقود (عناقيد)
Code	رمز
Combinatorial	تركيبي/ توافيقي
Compact	متراص
Comparator	مقارن
Competitive	تنافسي
Complement	مُتَمِّم
Commutative	تبديلي
Computer vision	رؤية حاسوبية
Concentration	تركيز
Conditional	شَرطيّ/ مشروط
Conditional connective	رابطة شرطية
Conservative	محافظ
Constraint	قَيدْ ـــ شرط مقيّد
Content-addressable memory	ذاكرة معنونة بالمحَتوى
Context layer	طبقة السياق

Continuous	مستمر
Contradiction	تناقض
Control	تحكم/ضبْط
Controller	مُتحكّم
Convergence	تقارب
Converter	مُحوّل/ مبدل
Correlation	ترابط
Coulomb	كولومب
Counterpropagation	الانتشار المتعاكس
Column	عمود
Covariance	تباين مشترك
Credit-assignment	تخصيص الاعتماد
Crossover	تقاطع/ تصالب
Cumulative	تراكمي

-D-

Data compression	ضغط المعطيات
Decision region	منطقة القرار
Decoder	مُفكّك ترميز
Definite integral	تكامل محدود
Dendrite	فرع شجري
Derivative	مشتق
Detection	كشف
Diagonal matrix	مصفوفة قطرية
Diagnosis	تشخيص
Diagram	مخطط
Differential equation	معادلة تفاضلية
Dilation	تمدد
Dimension	بُعْد
Discrete	مُقطّع/ منفصل
Dissipative	مبدد
Distribution	توزيع

Distributive	توزيعي
Disturbance	اضطراب

-E-

Echo cancellation	حذف الصدى
Eigenvalue	قيمة خاصة
Eigenvector	شعاع خاص
Electrocardiogram	مخطط، تخطيط القلب
Electrochemical	إلكتروكيميائية
Electrode	إلكترود
Electronic switching	تبديل إلكتروني
Element	عنصر
Embedding	تضمين
Empty set	مجموعة خالية
Emulation	محاكاة ـــ مضاهاة/ تقليد
Encoding	ترميز
Encoder	مُرمِّز
Energy	طاقة
Entropy	أنتروبي
Epoch	دور
Equalization	موازنة ـــ تسوية
Equation	معادلة
Estimation	تقدير
Estimator	مُقدِّر
Erasing	المحو ـــ الحذف
Euclidean	الإقليدي
Excitatory	مُهيِّج/ مُحرِّض
Expectation value	قيمة متوقعة
Expert system	نظام خبير
Exponent	أُسّ
Expression	تعبير
Extrapolation	استيفاء خارجي

-F-

Far-end	النهاية البعيدة
Fault	خطأ/ عطل/ خلل
Feature	سمة/ ميزَة
Feature maps	خرائط سمات
Feedback	تغذية راجعة
Feedforward	تغذية أمامية
Filtering	ترشيح
Fitting	إلباس
Flow	تدفق انسياب/ جريان
Forecasting	تنبؤ
Format	مَصَاغَة
Function	وظيفة/ إجراء
Fusion	دمج
Fuzzy	ترجيجي ـــ عائم ـــ مبهم

-G-

Generalization	تعميم
Genetic algorithm	خوارزمية جينية
Global minimum	أصغر شامل/ شمولي
Gradient descent	تدرج الهبوط
Grid	شَبَكْ
Grouping	تجمّع

-H-

Heteroassociative	ترافق مغاير
Heuristic method	طريقة تجريبية كسبية
Hidden layer	طبقة مخفية
Histogram	مخطط [بياني] نسيجي
Hyberbolic	زائدي المقطع
Hyberbolic tangent	ظل قطعي
Hybrid	هجين

-I-

Idempotent	اللانمو
Identification	تعريف
Identifier	مُعرِّف/ مُعيِّن [الهوية]
Identity matrix	مصفوفة واحدية
Identity function	تابع التماثل
Image processing	معالجة الصور
Implication	التضمين
Indefinite integral	تكامل غير محدود
Index	فهرس/ دليل
Inequality	متراجحة
Inference	استدلال
Information theory	نظرية المعلومات
Inhibitory	مُخمِّد
Initial value	قيمة أولية
Input	دخل
Integral	تكامل
Intensity	شدة
Interpolation	استيفاء داخلي- توليد
Interval	مجال
Invariant	لاتغيري
Inversion	قلب/ عكس

-J-

Joint	مشترك

-K-

Kinetic energy	طاقة حركية

-L-

Layer	طبقة
Learning	تعليم
Least squares	مربعات صغرى
Limit	نهاية/ حد

Linear	خطي
Linear independent	مستقل خطياً
Local minimum	أصغر محلي
Long term	أجل طويل

-M-

Manifold	جملة مولدة
Mapper	مطبق
Mapping	تطبيق /مقابلة/ إسقاط [طباقياً]
Marginal	هامشي
Mating	تزاوج
Matrix	مصفوفة
Maximization	تعظيم/ تكبير
Mean	متوسط/ وسطي
Mean square error	متوسط مربع الخطأ
Median	وسط، الفاصل في الوسط
Membrane potential	كمون الغشاء
Memorization	تذكر (وضع في الذاكرة)
Metric	مسافة
Mode	نمط
Model	نموذج
Modeling	نمذجة
MODEM	مودم: تعديل/فك تعديل
Moment	عزم
Momentum	كمية حركة
Monitoring	مراقبة
Multilayer	متعدد الطبقات
Multisensor	متعدد الحساسات
Mutation	طفرة وراثية

-N-

Near-end	النهاية القريبة
Neuro-fuzzy	عصبوني عائم

Neuro-logic	منطق عصبوني
Neuron	عصبون
Network	شبكة
Node	عقدة
Noise	ضجيج
Nonlinear	لاخطي
Nonparametric	موسطات
Norm	نظيم
Normal distribution	توزيع نظامي
Normalization	استنظام
Nucleus	نواة

-O-

Object	غرض
Off-line learning	تعليم مفصول (مؤجل) غير مباشر
Offspring	نسل
On-line	موصول [إلى الخط] مُتاح
Optimization	استمثال
Orbit	مدار
Orthogonal	متعامد
Outer product	جداء خارجي
Output	خرج

-P-

Parallel processing	معالجة متوازية
Parameter	مُوسط
Parity	ندِّية، ندية
Pattern	نَموذج- شكل
Pattern recognition	تعرّف الأشكال [النمطية]
Percentile	نسبة مئوية
Perception	إدراك
Performance	أداء – إنجاز
Phase	طور

Photoreceptor	مستقبل ضوئي
Pixel	بكسل (عنصر صورة)
Plasticity	اللدونة
Principal component	مركبة أساسية
Pocket	محفظة
Polynomial	كثير حدود
Prediction	تنبؤ
Probability	احتمال
Process	اجرائية
Processing	معالجة
Prototypical	نمذجة أولية
Pseudo-inverse	معكوس مُفترض

-Q-

Quantization	استكمام
Quantizer	مكمم
Quickprop	الانتشار السريع

-R-

Radial	شعاعي
Random	عشوائي
Random variable	متحول عشوائي
Range	مجال
Rank	رتبة
Rate	معدل
Recurrent	تكراري
Regression	انكفاء
Reinforced	معززة
Reinforcement	تقوية–تعزيز
Relative	نسبي
Relaxation	استرخاء
Representation	تمثيل
Resonance	طنين

Response	استجابة
Retrieval	استحضار
Risk	مجازفة
Robust	متين/ منيع (حصين)
Row	سطر
Rule	قاعدة (ناظمة)

-S-

Scalar product	جداء سُلّمي
Schedule	يُجدول (زمنياً) [جدول زمني]
Segment	مُقْتَطَع
Selection	تحديد/ انتقاء
Self-adaptive	متكيف ذاتياً
Self-growing	نمو ذاتي
Sensor	مُحس
Servo	[محَرك] تخديم
Set	مجموعة
Shift register	سجل إزاحة
Short term	أجل قصير
Signal	إشارة
Simulated annealing	محاكاة التلدين
Simulated networks	شبكات المحاكاة
Simulation	محاكاة
Soft computing	حساب لين
Soma	جسم الخلية
Spectrum	طيف
Stability	الاستقرار
Stable	مُستقر
Standard deviation	انحراف معياري/انحراف قياسي
State	حالة
Statistical	إحصائي
Stimulus	منبه

Stochastic	عشوائي
Storage	وسيطه خزن (خزّان)
Structure	بنية
Supervised −learning	تعليم بمعلم
Synapse	ليف
Synchronous	متزامن
System	نظام

-T-

Target output	الخرج الهدف (أو المنشود)
Term	حد−أجل
Theory	نظرية
Threshold	عتبة
Tolerance	تسامح
Toutology	كامل
Training	تدريب
Trajectory	مسار
Transformation	تحويل
Transformer	مُحوِّل
Transitive	متعدية
Transpose	منقول
Truth table	جدول الحقيقة

-U-

Unambiguous	غير غامضة
Uncertainty	شك
Unclamped mode	نمط عدم الإلزام
Unit	وحدة
Universal set	مجموعة عميمة
Unstable	غير مستقر

-V-

Valid argument	محاكمة صحيحة
Value	قيمة

Variance	تباين
Vector	متجه
Vigilance	يقظة–احتراس
Voxel	عنصر حجم

-W-

Weight	وزن/ ثقَل
Weighted sum	مجموع مُثقّل

المراجـــع

[1]- Hebb, D. O. *"The Organization of Behavior"* Wiley, New York, 1949.

[2]- Hopfield, J. J. & Tank, D. W. *"Neural Computation of Decisions in Optimization Problems"* Biological Cybernetics, Vol.52, pp.141-152, 1985.

[3]- Patterson, D. W., Chan, K. H., Tan, C. M. *"Time Series Forecasting with Neural Networks: A Comparative Study"* Proceedings of the International Conference on Neural Networks, Applications of signal processing (NNASP-93), Singapore, pp.269-274, 1993.

[4]- Zak, M. *"An Unpredictable - Dynamics Approach to Neural Intelligence"* IEEE Expert, pp.4-10, August, 1991.

[5]- Lippman, R. P. *"An Introduction to Computing with Neural Nets"* IEEE ASSP Magazine, pp.4-22, April, 1987.

[6]- Linsker, R. *"Self-Organization in a Perceptual Network"* IEEE Computer, Vol.21, No.3, pp. 105-117, 1988.

[7]- Kohonen, T. *"Self-Organized Formation of Topologically Correct FeatureMaps"* Biological Cybernetics, Vol.43, pp.59- 69, 1982.

[8]- Fukushima, K. & Miyake, S. *"Necognitron:A New Algorithm for Pattern Recognition Tolerant of Deformation and Shifts in Position"* Pattern Recognition, Vol.15, No.6, pp. 455 - 469, 1982.

[9]- Grossberg, S. *"Nonlinear Neural Networks: Principles, Mechanisms, and Architectures"* Neural Networks, Vol.1, No.1, pp. 17-61, 1988.

[10]- Hopfield, J. J. *"Neural Networks and Physical Systems with EmergentCollective Computational Abilities"* Proceedings of the National Academy of sience, Vol.79, pp. 2554-2558, 1982.

[11]- Anderson , J. A. *"Cognitive and Psychological Computation with Neural Models"* IEEE Transactions on Systems, Man, and Cybernetics, Vol. SMC-13, No.5, pp.799 - 815, 1983.

[12]- Kosko, Bart *"Adaptive Bidirectional Associative Memories"* Applied Optics, Vol.26, No.23, pp.4947-4960, 1987.

[13]- Shannon, C. E. *"A Mathematical Theory of Communication"* Bell System Technical Journal, Vol.27, pp.379-423, pp.623 - 656, 1948.

[14]- Gutzwiller, M. C. *"Quantum Chaos"* Scientific American, pp.78-84, January, 1992.

[15]- Lorentz, E. N. *"Computational Chaos-A Prelude to Computational Instability"* Physica D, Vol.35, pp.299-317, 1989.

[16]- Takens, F. *"Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence"* Warwick, 1980, Lecture Notes in mathematics No.898,Rand, D. and Young, L. S., eds., Springer, Berlin, pp.366-381,1981.

[17]- Mane, R. *"Dynamical Systems and Turbulence"* Warwick, 1980, Lecture Notes in mathematics No.898, Rand, D. and Young, L. S., eds., Springer, Berlin, pp.230-242, 1981.

[18]- Ababarbanel, H. D. I. & Reggie Browen, & Tsimring, L. S. *"The Analysis of Observed Chaotic Data in Physical Systems"* unpublished document preprint,1993.

[19]- McCulloch, W. S. & Pitts W. *"A Logical Calculus of the Ideas Immanent in Nervous Activity"* Bulletin of mathematical Biophysics, Vol.5, pp.115-133, 1943.

[20]- Rumelhart, D. & McClelland, J. *"Parallel Distributed Processing"* Vol.1, eds., MIT press, Cambridge, MA, 1986.

[21]- Rosenblatt, F. *"The Perceptron: Probabilistic Model for Information Storage and Organization in the brain"* Psychological Review, Vol. 65, pp.386-408. Reprinted in Anderson & Rosenfeld [1988], pp.92-114, 1958.

[22]- Gallant, S. I. *"Neural Network Learning and Expert Systems"* MIT Press, Cambridge, MA, 1993.

[23]- Hertz, J. A. & Krogh, A. & Palmer, R. G. *"Introduction to the Theory of Neural Computation"* Addison-Wesley, Redwood City, CA, 1991.

[24]- Minsky, M. L. & Papert, S. A. *"Perceptrons"* Expanded edition, Cambridge, MIT Press, MA, 1988.

[25]- Arbib,M. A. *"Brains, Machines, and Mathematics"* 2^{th} ed., Springer-Verlag, New York, 1987.

[26]- Rosenblatt, F. *"Principles of Neurodynamics"* New York, Spartan, 1962.

[27]- Burr, D. J. "*An Improved Elastic Net Method for the Traveling Salesman Problem*" Proceedings of the IEEE International Conference on Neural Networks, San Diago, Vol.1, pp.69-76, 1988.

[28]- Patterson, D. W. "*Artificial Neural Networks: Theory and Applications*" Prentice Hall, Simon & Schuster(Asia), Singapore,1996.

[29]- Widrow, B. & Hoff, M. E. "*Adaptive Switching Circuits*" IRE WESCON Convention Record, New York, 1960.

[30]- Kohonen, T. "*Associative Memory :A System Theoretic Approach*" Springer, New York, 1977.

[31]- Stone, G. "*Parallel Distributed Processing*" Vol.1, MIT Press, MA, Cambridge, 1986.

[32]- Widrow, B. & Winter, R. "*Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition*" IEEE Computer, Vol.21, No.3, pp.25-39, 1988.

[33]- Andes, D. & Widrow, B. & Leher, M. & Wan, E. "*MRIII: A Robust Algorithm for Training Analog Neural Networks*" Proceedings of the International Joint Conference on Neural Networks, Seattle, WA, Vol.I, Erlbaum, Hillsdale, NJ, pp.533-536, 1991.

[34]- Widrow, B. & Winter, R. G. & Baxter, R. A. "*Learning Phenomena in Layered Neural Networks*" Proceedings of the First IEEE International Conference on Neural Network, San Diago, 1987.

[35]- Widrow, B. & Rumelhart, D. & Lehr, M. "*Neural Networks : Applications in Industry, Business and Science*" Communication of the ACM, Vol.37, No.3, pp.93-105, 1994.

[36]- Tesauro, G. "*Simple Neural Models of Classical Conditioning*" Biological Cybernetics, Vol.55, pp.187-200, 1986.

[37]- Szu, H. H. "*Neural Networks: Theory, Applications and Computing*" Lecture Notes for UCLA Engineering Short Course, Engineering 819.185, March 20-23, 1989.

[38]- Anderson, J. A. "*A Simple Neural Network Generating an Interactive Memory*" Mathematical Biosciences, Vol.14, pp.197-220. Reprinted in Anderson & Rosenfeld [1988], pp.181-192, 1972.

[39]- Hopfield, J. J. "*Neurons with Graded response Have Collective Computational Properties Like Those of two-State Neurons*" Proceedings of the National Academy of science, Vol.81, pp.3088-3092, 1984.

[40]- Cohen, M. Y. & Grossberg, S. "*Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks*" IEEE Transectios on Systems, Man and Cybernetics, Vol.13, pp.815-826, 1983.

[41]- Bilbro, G. & Miller, T. K. & Snyder, W. E.& Van den Bout, D. E. & White, M. "*Optimization by Men Field Annealing, in Advances in Neural Information Processing Systems*" T.D.S. Touretaky, Morgan Kaufmann, San Mateo, CA, pp.91-98, 1988.

[42]- Kosko, Bart "*Bidirectional Associative Memories*" IEEE transactions on systems, Man and Cybernetics, Vol.SMC-18,pp.49-60, 1988-a.

[43]- Kosko, Bart "*Feedback Stability and Unsupervised Learning*" Proceedings of the IEEE International Conference on Neural Networks, San diago,Vol.1, pp.141-149, 1988-b

[44]- Abu-Mostafa, Y. S. & St Jacques, J. –M. "*Information Capacity of the Hopfield Model*" IEEE transactions on Information Rheory, IT-31, pp.461-464, 1985.

[45]- McEliece, R. J. & Posner, E. C & Rodemich, E. R. & Venkatesh, S. S. "*The Capacity of the Hopfield Associative Memory*" IEEE Transactions on Information Theory, IT-33, pp.461-482, 1987.

[46]- Hecht-Nielsen, R. "*Neurocomputing*" Readings, MA, Addison-Wesley, 1990.

[47]- Anderson, J. A. & Silverstein, J. W. & Ritz, S. A. & Jones, R. S. "*Distictive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model*" Psychological Review, Vol.84, pp.413-451, 1977.

[48]- Hecht-Nielsen, R. "*Applications of Counterpropagation Networks*" Neural networks,Vol. 1(2), pp.131-139, 1988.

[49]- Kosko, B."*Bidirectional Associative Memories*" IEEE Transactions on Systems, Man, and Cybernetics, Vol.18, pp.49-60. Reprinted in Anderson, Pellionize & Rosenfeld[1990], 1988.

[50]- Rosenblatt, F. "*Principles of neurodynamics: Perceptrons and the Theory of Brain Mechanisms*" Spartan Books, Washington, DC, 1961.

[51]- Werbos, P. J. *"Beyond Regression: New Tools for prediction and Analysis in the Behavioral sciences"* PH.D. Thesis, Harvard University, 1974.

[52]- Parker, D. B. *"learning Logic"* Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA, 1985.

[53]- Rumelhart, D. J. & Zipser, D. *"Feature Discovery by Competitive Learning"* Cognitive Science, Vol.9, pp.75-112, 1985.

[54]- Widrow, B. & Stearns, S. D. *"Adaptive Signal Processing"* Prentice-Hall, Englewood Cliffs, NJ, 1985.

[55]- Cottrell, G. W. & Munro, P. & Zipser, D. *"Image Compression by Back Propagation: An Example of Extensional Programming"* In N. E. Sharkey, ed., Models of Cognition: A Review of Cognitive Science, Norwood, NJ: Ablex Publishing Corp, pp.208-240, 1989.

[56]- Rumelhart, D. E. & Hinton, G. E. 7 Williams, R. J. *"Learning Internal Representation by Error propagation, in Parallel Distributed Processing: Explorations in the Microstructures of Cognition"* Vol.1, Foundations, Rumelhart, D. E. & McClelland, MIT Press, Cambridge, MA, 1986.

[57]- Arozullah, M. & Namphol, A. *"A Data Compression System Using Neural Networks; Based Architecture"* International Joint Conference on Neural Networks, San Diago , CA, I: 531-536, 1990.

[58]- Sonehara, N. & Kawato, M. & Nakane, K. *"Image Data Compression Using a Neural Network Model"* International Joint Conference on Neural Networks, Washington, DC,Vol. II, pp.35-41, 1989.

[59]- Yu, X. H. *"On the Nonexistence of Local Minima of the Backpropagation Error Surfaces"* Proceedings of the IJCNN-91, Singapore, Vol.2, pp.1272-1277, 1991.

[60]- Jacobs, R. A. *"Increased Rates of Convergence Through Learning Rate Adaptation"* Neural Networks, Vol.1, No.4, pp.295-307, 1988.

[61]- Fahlman, S. E. *"An Empirical Study of Learning Speed in Backpropagation Networks"* Carnegie Mellon Report No.CMU-CS-88-162, 1988.

[62]- Hirose, Yoshio, Koichi Yamashita & Shimpei Hajiya *"Backpropagation Algorithm Which Varies the Number of Hidden Units"* Neural Networks, Vol.4, No. 1, pp.61-66, 1991.

[63]- Chen, C. L. 7 Nutter, R. S. *"Improving the training Speed of Three-Layer Feedforward Nets by Optimal Estimation of the Initial Weights"* Proceedings of the IJCNN-91, Singapore, Vol.3, pp.2063-2038, 1991.

[64]- Denueux, T. & Lengelle, R. & Canu, S. *"Initialization of Weights in a Feedforward Neural network Using Prototypes"* Proceedings of te ICANN-91, Espoo, Finland, pp.623-628, 1991.

[65]- Kim, Y. K. & Ra, J. B. *"Weight Value Initialization for Improving training Speed in the Backpropagation Network"* proceedings of the IJCNN-91, Singapore,Vol.3, pp.2396-2401, 1991.

[66]- Haario, H. & Jokinen, P. *"Increasing the Learning Speed of Backpropagation Algorithm by Linearization, in Artificial Neural Networks"* Proceedings of the ICANN-91, Espoo, Finland, Vol.1, Kohonen, T. & Makisara, K. & Simula, O. & Kangas, J., North-Holland, Amesterdam, pp.629-634, 1991.

[67]- Tollenaere, Tom *"SuperSAB: Fast Adaptive Backpropagation with Good Scaling Properties"* Neural Networks, Vol.3, pp.561-573, 1990.

[68]- Minai, A. A. & Williams, R. D. *"Backpropagation Heuristics: A study of the Extended Delta-Bar-Delta Algorithm"*Proceedings of the IJCNN-90, San diago, Vol.1, pp.595-600, 1990.

[69]- Sato, A. *"An Analytical Study of the Momentum Term in a Backpropagation Algorithm"* Proceedings of the ICANN-91, Espoo, Finland, pp.617-622, 1991.

[70]- Fogelman-Soulie, F; *"Neural Network Architectures and Algorithms: A perspective, Artificial Neural Networks"* Vol.1, Kohonen, T. & Simula, O. & Kangas, J., North-Holland, Amsterdam, pp.605-615, 1991.

[71]- Matsuoka, K. & Yi, J. *"Backpropagation Based on the Logarithmic Error Function and Estimation of Local Mimima"* Proceedings of the IJCNN-91, Singapore, Vol.2, pp.1117-1122, 1991.

[72]- Van ooyen, A. & Nienhuis, B. "Improving the Convergence of the Backpropagation Algorithm" Neural Networks, Vol.5, pp.465-471, 1992.

[73]- Battiti, R. *"First-and Second-Order Methods for Learning : Between Steepest Descent and Newton's Method"* Neural Cpmputation,Vol.4, No.2, pp. 141-166,1992.

[74]- Bishop, C. "*Exact Calculation of the Hessian Matrix for the Multilayer Perceptron*" Neural Computation,Vol.4,pp.491-501,1992.

[75]- Ballard, D. H. & Brown, C. M. "*Computer Vision*" Prentive-Hall, Englewood Cliffs, NJ, 1982.

[76]- Thorpe, C. & Hebert, M. & Kanade, T. & Shafer, S. "*Toward Autonomous Driving: The CMU NavLab*" IEEE Expert, pp.31-42, August, 1991.

[77]- Kanade, T. & Reed, M. L. & Weiss, L. E. "*New Technologies and Applications in Robotics*" Communications of the ACM, Vol.37, No.3, pp.58-67, 1994.

[78]- Staib, W. E. "*The Intelligent Arc Furance: Neural Networks Revolutionize Steel-Making*" Proceedings of INNS meeting, World Congress on Neural Networks, Portland, Or, Vol.I, pp.466 - 469, 1993.

[79]- Asakawa, Kazuo & Hideyuki Takagi "*Neural Networks in Japan*" Communications of the ACM,Vol.37, No.3, pp.106 - 12, 1994.

[80]- Rock, D. & Malkoff D. & Stewart, R. "*AI and Aircraft Health Monitoring*" AI Expert, pp.28-35, February, 1993.

[81]- Morose, R. A. "*A Financial Neural Network Application*" AI Expert, pp.50-53,May, 1990.

[82]- Morose, R. A."*A Financial Neural Network Application, in Neural Networks in Finance and Investing*" Trippi, R. R. & Turban, E., Probus Publishing, Chicago, pp.75-83, 1993.

[83]- le Cun. Y."*Models Connexionnistes de l'apprentissage*" Doctoral Dissertation, University of Pierre and Marie Curie, Paris, 1987.

[84]- le Cun, Y. & Boser, B. & Denker, J. S. & Solla, S. & Howard, R. & Jackel, L. "*Backpropagation Applied to Handwritten Zipcod Recognition*" Neural Computation, Vol.1, pp.541-551, 1990.

[85]- Keeler, J. D. & Rumelhart, D. E. & Leow, W. K. "*Integrated Segmentation and Recognition of Hand-Printed Numerals, in Neural Information Processing Systems*" Vol.3, Lippman, R. P. & Moody, J. E. & Touretzky, D. S., Morgan Kaufmann, San Mateo, CA, pp.557-563, 1992.

[86]- Lee, C. M. & Patterson, D. W. "*Occluded Object Recognition: An Approach Which Combines Neurocomputing and Conventional Algorithm*" Proceedings of the IJCNN-91, Singapore, Vol.2, pp.2612-2617, 1991.

[87]- Yanicoglu, B. A. & Sandon, P. A. *"off-Line Cursive Handwriting recognition Using Neural Networks"* Proceedings of the SPIE Applications of Artificial Neural Network IV, Orlando, FL, pp.102-106, 1993.

[88]- Hamilton & Hufnagel *"Early Detection of Epileptic Attacks, in Applications of neural Networks"* Schuster, H. G., VCH Verlagsgesellschaft, Weinheim, pp.173-178, 1992.

[89]- Colombi, J. M. & Anderson, T. R. & Rogers, S. K. *"Auditory Model Representation for Speaker Recognition"* Proceedings of the SPIE Applications of Artificial neural Networks IV,Orlando,FL,pp.9-14,1993.

[90]- Minsky, M. L. & Papert, S. A. *"Perceptrons"* MA:MIT Press,Original edition, 1969.

[91]- Pinda, F. J. *"Generalization of Backpropagation to Recurrent Neural Networks"* Physical Review Letters,Vol.59, pp.2229-2232, 1987.

[92]- Pinda, F. J. *"Dynamics and Architecture for Neural Computation"* Journal of Complexity,Vol.4, pp.216-245, 1988.

[93]- Pinda, F. J. "Recurrent Backpropagation and the Dynamical Approach to Adaptive Neural Computation" Neural Computation, Vol.1, pp.161-172, 1989.

[94]- Almeida, L. B. *"A Learning Rule for Asynchronous Perceptrons with Feedback in a Networks"* San Diago, Vol.2, pp.609-618, 1987.

[95]- Williams, R. J. & Zipser, D. *"A learning Algorithm for Continually Running Fully Recurrent Neural Networks"* Neural Computation, Vol.1, pp.270-280, 1989.

[96]- Pearlmutter, B. A. *"Dynamic Recurrent Neural Networks"* Report CMU-CS-88-91,School of Computer Science, Carnegir Mellon University, Pittsburgh, PA, 1988.

[97]- Zipser, D. *"A Subgroubing strategy that Reduces Complexity and speeds Up Learning in recurrent Networks"* Neural Computation, Vol.1, pp.552-558, 1989.

[98]- Williams, R. J. & Peng, J. *"An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Traj*ectories" Neural Computation, Vol.1, pp.270-278, 1989.

[99]- Atiya, A. F. *"Learning on a General Network, in Neural Information Processing Systems"* Anderson, D. Z., American Institute of Physics, New York, 1988.

[100]- Rumelhart, D. E. & Hinton, G. E. & Williams, R. J. "Learning Internal Representations by Error propagation" In Rumelhart, D. E. & McClelland,J. L.,eds., *Prallel Distributed Processing*, Vol.1, Chpter 8. Reprinted in Anderson & Rosenfeld [1988], pp.675 - 695, 1986 -a.

[101]- Rumelhart, D. E. & Hinton, G. E. & Williams, R. J. *"Learning Representations by Error propagation"* Nature, Vol.323, pp.533-536. Reprinted in Anderson & Rosenfeld[1988], pp.696-699, 1986-b.

[102]- Elman, J. L. *"Distributed representations, Simple Recurrent networks ,and Grammatical Structure, Machine Learning"* Vol.7, pp. 195- 225,1991.

[103]- Servan-Schreiber, D; & Cleeremans, A. & McClelland, J. L. *"Graded State Machines: The representation of temporal Contingencies in Simple recurrent networks"* Machine Learning, Vol.7, pp.161-193, 1991.

[104]- Sterzing, V. & Schurmann, B. *"Recurrent Neural Networks for Temporal Learning of Time Series"* Proceedings of the IEEE International Comference on Neural Networks, San Francisco, Vol.2, pp. 843-846, 1993.

[105]- Li, Liang & Haykin, S. *"A Cascade Recurrent Neural Network for Real-Time Nonlinear Adaptive Filtering"* Proceedings of the IEEE International Conference on Neural Networks, San Francisco, pp.857-862,1993.

[106]- Mori, H. & Ogasawara, T. *"A Recurrent Neural Network Approach to Short-Term Load Forecasting in Electronic Power Systems"* Proceedings of the World Congress on Neural Networks, Portland, OR, Vol.I, pp.342-345, 1993.

[107]- Rao, S. S. & Rammamurt, V. *"A Hybrid Technique to Enhance the Performance of Recurrent Neural Networks for Time Series Prediction"* Proceedings of the IEEE International Conference on Neural Networks, San Francisco, 1993.

[108]- Freisleben, B. *"The Composer: A Network for Musical Appications, in Artificial Neural Networks"* Vol.2, Aleksander, I. & Taylor, J., eds., North-Holland, Amesterdam, pp.1663-1666, 1992.

[109]- Fernande, S. & Islam, F. & Utama, P. & Watson, K. *"High Impedance Fault Detection Using recurrent Network, Artificial neural Networks"* Vol.2,

Aleksander, I. & Taylor, J., eds., North-Holland, Amsterdam, pp.1615-1618, 1992.

[110]- Hoshino, T & Kano, M. & Endo, T. *"Optical Control with a Recurrent Network and a priori Knowledge of the System"* Proceedings of the IJCNN-91, Singapore, pp.226-231, 1991.

[111]- Imai, K. *"Simple Recurrent Neural Networks Applied to the Recognition of a Lateral String of Letters"* Private communication between l'auther and Reference[18], unpublished paper, 1991.

[112]- Wang, P. Z. "Truth-Valued Flow Inference Theory and its Application, in Advances in Fuzzy Systems: Application and Theory" Wang, P. Z. & Loe, K. F., eds., World Scientific, Singapore, 1993.

[113]- Hinton, G. E. & Sejnowski, T. J. "Analyzing Cooperative Computation" Proceedings of the Fifth Annual Conference of the Cognitive Science Society, Rochester, NY, pp.448-453, 1989.

[114]- Ackley, D. H. & Hinton, G. E. & Sejnowski. T. J. *"A Learning Algorithm for Boltzmann Machines"* Cognitive Sience,Vol.9,pp.147-69,1985.

[115]- Fausett, L. *"Fundamentals of Neural Networks: Architecturesm Algorithm, Applications"* Prentice Hall, Englewood Cliffs, NJ,1994.

[116]- Gemen, S. & Gemen, D. *"Stochastic Relaxtation, Gibbs Distribhtions and the Bayesian Restoration of Imaes"* IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, pp.721-741, 1984.

[117]- Szu, H. & Hartly, E. Physics Letters, pp.157-162, 1987.

[118]- Garey, M. & Johnson, D. *"Computers and Intractability: A Guide to the Theory of NP-Completeness"* Freeman, W. H., San Francisco, 1979.

[119]- Light, L. W. & Anderson, P. *"Designing Better Keyboards via Simulated Annealing"* AI Expert, pp.20-7,September, 1993.

[120]- Aarts, E. & Korst , J. *"Simulated Annealing and Boltzmann machines : A Stochastic Approach to Combinatorial Optimization and Neural Computing"* Wiley, New York, 1989.

[121]- Lawler, E. L. & Lenstra, J. K. & Rinnooy Kan, A. H. G.& Shmoys, D. B. *"The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization"* New York, John Wiley & Sons, 1985.

[122]- Wilson, G. V. & Pawley, G. S. *"On the stability of the Traveling Salesman Problem Alogrithm of Hopfield and Tank"* Biological Cybernetics, Vol.58, pp.63-70, 1988.

[123]- Szu, H. H. *"Fast TSP Algorithm Based on Binary Neuron Output and Analog Neuron Input Using the Zero-Diagonal Interconnect Matrix and necessary and Sufficient Constraints of the Permutation Matrix"* IEEE International Conference on Neural Networks, San Diago, CA, Vol.II, pp.259-265, 1988.

[124]- Hopfield, J. J. & Tank, D. W. *"Neural Computation of Decisions in Optimization Problems"* Biological Cybernetics, Vol.52, pp.141-152, 1985.

[125]- Takefuji, Y. *"Neural network Parallel Computer"* Boston: Kluwer Academic Publishers, 1992.

[126]- Croall, I. F. & Mason, J. P. (eds.) *"Industrial Applications of neural networks"* In project ANNIE Handbook, Springer-Verlag, Berlin, 1991.

[127]- Moallemi, C. *"classifying Cells for Cancer diagnosis Using Neural Networks"* IEEE Expert, December, pp.8-12, 1991.

[128]- Sone, Tadashi *"Using Distributed Neural Networks to Identify Faults in Switching Systems"* Proceedings of the International Workshop on Applications of Neural Networks to telecommunications, Alspector, j. & Goodman, R.& Brown, T. X., Lawrence Erlbaum associates Hillsdale, NJ, 1993.

[129]- Abe, Shigeo & Kayama, M. & Taenaga, H. *"synthesizing neural networks for Pattern Recognition"* Proceedings of the IJCNN-9I, Singapore, pp.1105-10, 1991.

[130]- Caudill, M. *"Neural Networks Primer"* San Francisco, Miller Freeman, 1989.

[131]- Abu-Mostafa, Y. S. *"Vapnik-Chervonenkis Dimension : Information Versus Complexity in Learning"* Neural Computation, Vol.1, pp312-7, 1989.

[132]- Zadeh, L. A. *"Fuzzy Logic, Neural Networks , Soft Computing"* Communications of the ACM,Vol.37, No.3, pp.77-84, 1994.

[133]- Jahne, B. *"Digital Image Processing"* Springer-Verlag, Berlin, Germany, 1997

[134]- Ackley, D. H. "*A Connectionist Machine for Genetic Hillclimbing* " Kluwer Academic Publishers, Boston, 1987.

[135]- Barhin,J. & Gulati, S. & Zak, M. "*Neural learning of constrained Nonlinear Transformation*" Computer, Vol.22(6), pp.67-76, 1989.

[136]- Youssef, H. M. "Comparison of Neural Networks in Nonlinear System Modeling" proceedings of the World Congress on Neural Networks, Portland, OR, pp.IV5-9, 1993.

[137]- Weigend, A. S. & Gershenfeld, N. A. "*Time Series Prediction: Forecasting the Future and Understanding the Past*" Addition-Wesley, Reading, MA, 1994.

[138]- Carpenter,G. A. & Grossberg, S. "*The Art of Adaptive Pattern Recognition by a Self-Organizing Neural Network*" Computer,Vol.21, pp.77-88, 1988.

[139]- Carpenter,G. A. & Grossberg, S.& Rosen, D. B. "*Fuzzy Art: Fast Stable Learning and Categorization of Analog Input Patterns by an Adaptive Resonance System*" Neural Networks, Vol.4, pp.759-71, 1991.

[140]- Ahmad, S. & Tesauro, G. "*Scaling an Generalization in Neural networks*" In D. S. Touretzky, ed., Advances in Neural Information Processing Systems1, San Mateo, CA: Morgan Kaufmann, pp.160-168, 1989.

[141]- Szu, H. & Liu, K. W. & Chao, C. C. & Lin, K. F. & Hsu, K. T. & Medsker, L. Proceedings of the IJCNN-92, Beijing, pp.I-333,I-339, 1992.

[142]- Vemuri, V. "*Artificial Neural Networks: Theoretical Concepts*" Washington, DC: IEEE Computer Society Press, 1988.

[143]- Akiyama, Y. & Yamashita, A. & Kajiura, M. & Aiso, H. "*Combinatorial Optimization with Gaussian Machines*" International Joint Conference on Neural networks, Washington, DC, I:533-540, 1989.

[144]- Ritter, H. J. & Martinetz, T. & Schulten, K. J. "*Neural Computation and Self-Organizing Maps: An Introduction*" Addison-Wesley, Readings, MA, 1992.

[145]- Specht, D. F. "*Probabilistic Neural Networks*" Neural networks,Vol.3, pp.109-18,1990.

[146]- Xu, L. & Qja, E. & Susen, G. S."*Modified Hebbian Learning for Curve and Surface fitting*" Neural Networks, Vol.5(3), pp.441-457, 1992.

[147]- Yager, R. R. *"Modeling and Formulating Fuzzy Knowledge Bases Using Neural Networks"* Neural Networks,Vol. 7, No.8, pp.1273- 83, 1994.
[148]- Alman, W. f. *"Apprentices of Wonder: Inside the Neural Network Revolution"* New York, Bantam Books, 1989.
[149]- Baum, E. B. & Hausler, D. *"What Size Net Gives Valid Generalization?"* Neural Computation,Vol.1, pp.151-60, 1989.
[150]- Block, H. D. *"The Perceptron : A Model for Brain Functioning"* Reviews of Modern Physics, Vol.34, pp.123-35, 1962.
[151]- Blum, E. B."A Proposal for More Powerful Learning Algorithms" Neural Computation,Vol.1, pp.201-207, 1989.
[152]- Rade, L. & Westergren, B. *"Beta Mathematics Handbook"* Studentlitterratur, Chartwell0Bratt Ltd., 1990.
[153]- Caudill. M. & Butler, C. *"Naturally Intelligent Systems"* Cambridge, MIT Press, MA, 1990.
[154]- Dahl,E. D. *"Accelerated learning Using the Generalized Delta rule"* Proceedings of the First IEEE International Conference on Neural networks, San diago, 1987.
[155]- Murat Tekalp, A. *"Digital Video Processing"* Prentice-Hall, Simon & Schuster, NJ, 1995.
[156]- Oja,E. *"Principal Components , Minor Components , and Linear Neural Networks"* Neural Networks, Vol.5(6), pp.927-935,1992.
[157]- Miller, W. T. & Sutton, R. S. & Werbos, P. J. *"Neural Networks for control"* Cambridge , MIT Press, eds., MA, 1990.
[158]- Murry,D. "Tuning Neural Networks with Genetic Algorithms" AI Expert, pp.27-32, June, 1994.
[159]- Almeida, L. B., *"Backpropagation in perceptrons with Feedback "*In R. Eckmiller, & CH. Von der Malsburg, eds., Neural Computers. Berlin, Springer-Verlag, pp.199-208, 1988.
[160]- Amari,S-I *"A Theory of Adaptive Pattern Classifiers "* EEE Transactions on Electronic Computers,Vol.EC, pp.299-307, 1967.
[161]- Chan, S. C. & Hsu, L. S. & Loe, K. F. & The, H. H. *"Neural Logic networks"* Internel Publication of the National University of Singapore, Singapore, pp.1-54.1991.

[162]- Chang, C. F. & Sheu, B. & Thomas, J. *"Multilayered Backpropagation Neural Networks for Financial Analysis"* Proceedings of the INNS Meeting, World Congress on Neural Networks , Portland, OR,Vol.I,pp.445-50,1993.

[163]- Amari, S-I & Fujita, N. & Shinomoto, S. *"Four Types of Learning Curves"* Neural Computation,Vol.4, pp.605-618, 1992.

[164]- Levine, D. S. *"Introduction to Neural and Cognitive Modeling"* Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

[165]- MacGregor, R. J. *"Neural and Brain Modeling"* San Diago, Acadenic Press, 1987.

[166]- McCoulloch, W. S. *"Embodiments of mind"* Cambridge, MIT Press, MA, 1988.

[167]- Barr,D. S. & Mani, G. *"Using Neural Nets to Manage Investments"* AI Expert,Vol.9, No.2, pp.16-21, 1994.

[168]- Kosko, B. *"Neural Networks and Fuzzy Systems; A dynamical Systems Approach to Machine Intelligence"* Englewood Cliffs, NJ, Prentice-Hall, eds., 1992-a.

[169]- Kosko,B. *"Neural networks for signal Processing"* Englewood Cliffs, NJ, Prentice-Hall, eds., 1992-b.

[170]- Lawrence, J. *"Data preparation for a neural Network"* Neural Network Special report, AI Expert, pp. 15-21, 1993.

[171]- Anderson , J. A. *"A memory Model Using Spatial Correlation Functions"* Kybernetik,Vol.5, pp.113-9, 1968.

[172]- Anderson , J. A. *"Two models for Memory Organization"* Mathematical Biosciences, Vol.8, pp.137-60, 1970.

[173]- Anderson , J. A. & Rosenfeld, E. *"Neurocomputing: Foundations of Research"* Cambridge, MIT press, MA, 1988.

[174]- Anderson , J. A. & Rosenfeld, E. *"Neurocomputing2: Directions for Research"* Cambridge, MIT press, MA, 1990.

[175]-Leondes, Cornelius T. *"Neural Network Systems, Techniques and Applications"* ACADEMIC PESS,New York, 1998

[176]- Angeniol, B. & De La Croix Vaubois, G. & Le Texier, J.-Y. *"Self-Organizing Feature Maps and the Traveling Salesman Problem"* Neural Networks,Vol.1, pp.289-93, 1988.

[177]- Kohonen, T. "*How to Make a Machine transcribe Speech?*" In Applications of neural Networks, H. G. Schuster(ed.), VCH Verlags-gesellschaft, Weinheim, 1992.

[178]- Kohonen, T. "*Self-Organization and Associative Memory*" Second edition,Springer-Verlag, Berlin-Heidelberg, 1987.

[179]- Kohonen, T. "*Neural Phonetic Typewriter*" IEEE Computer,Vol.21, No.3, pp.11-22, 1988.

[180]- Bornholdt, S. & Graudez, D. "*General Asymmetric Neural Networks and Structure Design by Genetic Algorithms*" Neural Networks, Vol.5, pp.327-34, 1992.

[181]- Chiuh,T. D. & Tang, T. T. & Chen, L. G. "*Vector Quantization Using Tree-Structured Self-Organizing Feature Maps*" Proceedings of the International Workshop on Applications of Neural Networks to Telecommunication, J. Alspector & R. Goodman & T. X.Browk(eds.), Lawrence Erlbaum Associstes, Hillsdale, NJ, pp.259-65, 1993.

[182]- Cohen, M. A. & Tesauro, G. "*How Tight are the Vapnik-Chervonenkis bounds?*" Neural Computation,Vol.4, pp.249-69, 1992.

[183]- DARPA "*DARPA neural Network Study*" Final Report, Cambridge, MA:Massachusetts Institute of technology, Lincoln Laboratory, 1988.

[184]- Fukushima, K. "*A Neural Network for Visual Pattern Recognition*" IEEE Computer,Vol. 21, No. 3, pp. 65-75, 1988.

[185]- Fukushima, K. & Wake, N. "*Handwritten Alphabetic Character Recognition by the Neocognitron*" IEEE Transactions on Neural Networks, Vol.2, No.3, pp.355-65, 1991.

[186]- Geman, S. E. & Bienenstock, E. & Doursat, R. "*Neural Networks and the Bias/Variance Dilemma*" Neural Computation,Vol.1, pp.1-58, 1992.

[187]- Burrascano, P. "*Learning Vector Quantization for the probabilistic Neural Network*" IEEE Transactions on Neural Networks, Vol.2, No.4, pp.458-61, 1991.

[188]- Dayhoff,J. E. "*Neural network Architectures*" New York, VanNostrand Reinholt, 1990.

[189]- Gruber, S. L. & Villalabos, L. & Olsson, J. "*Neural Networks for Webb-Process Inspection*" Proceedings of the SPIE Applications of Artificial neural Networks IV, pp.491-503, 1993.

[190]- Harston, C. T. "*Business with Neural networks*" In A. J. Maren & C. T. Harston & R. M. Papeds., Handbook of Neural Computing Applications. San Diago: Academic Press, pp391-400, 1990.

[191]- Drago, G. P. & Ridella, S. "*Cascade Correlation : An Incremental Tool for Function Approximation, in Neural Information Processing Systems*" Vol.2, R. P. Lippman, J. E. Moody & D. S. Touretzky, eds., Morgan Kaufmann, San Mateo, CA, pp.750-756, 1991.

[192]- Fausett, L. "*Fundamentals of Neural Networks*" Prentice-Hall, Englewood Cliffs, NJ, 1994.

[193]- Kasuba, T. "*Simplified Fuzzy ARTMAP*" AI Expert, Vol.8, No.11, pp.18-25, 1993

[194]- Goldberg, D. E. "*Genetic Algorithms*" Addison-Wesley, Reading, MA, 1989.

[195]- Hecht-Nielsen, R. "*Theory of the Backpropagation Neural Network*" International Joint Conference on Neural networks, Washington, DC, Vol. I, pp.593-605, 1989.

[196]- Freedman, D. H. "*Brainmakers*" Simon & Schuster, New York, 1994.

[197]- Grenender, U. "*Abstract Inference*" Wiley, New York, 1981.

[198]- Grossberg, S. "*Studies of Mind And brain*" Boston, Reidel, 1982.

[199]- Hopfield, J. J. & Tank, D. W. "*Computing with Neural circuits*" Science,Vol. 233, pp.625-633, 1986.

[200]- Karr, Chuck "*Applying Genetics to Fuzzy Logic*" AI Expert, pp. 38 - 43, March, 1991.

[201]- Cover, .M. T. & Thomas, J. A. "*Elements of Information Theory*" John Wiley & Sons,USA, 1991.

[202]- Reily, D. L. & Cooper, L. N. & Elbaum, C. "*A Neural Model for Category Leaning*" Biological Cybernetics, Vol.45,pp.34-51,1982.

[203]- Reily, D. L. & Cooper, L. N. & Elbaum, C. "*Learning System Architectures Composed on Multiple Learnin Modules*" Proceedings of the IEEE International Conference on Neural Networks,Vol.2,pp.495-503,1987.

[204]- Scofield, C. L. *"Learning Internal Representations in the Coulomb Energy Network"* IEEE International Conference on Neural Network, San Diego ,CA,Vol.I,pp.271-276,1988.

[205]- Fahlman, S. E. & Labiere, C. *"The Cascace-Correlation Learning Architecture"* Carnegie MillonReport,No. CMU-CS-88-162,1990.

[206]- Littman,E. & Ritter, H. *"Cascade Network Architectures"* Proceedings of the International Joint Conference on Neural Networks, Baltimore, Vol. II, pp. 398 - 404 , 1992.

[207]- Drago, G. P. & Ridella, S. *"An Optimum Weights Initializations for Improving Scaling Relationships in BP learning in Artificial Neural Networks"* Vol.1, Kohonen, T. & Makisara, O. Simula & Kangas,J., eds., North - Holland, Amsterdam, pp.1519 - 1522,1991.

[208]- Gallant, S. I. *"Connectionist Learning Algorithm with Provable Generalization and Scaling Bounds"* Neural Networks,Vol.3, pp . 191-201,1990.

[209]- Frean, M. *"The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks"* Neural Computation,Vol.2, pp. 198 – 209 , 1990.

[210]- Li, Wei & Nasrabadi, M. *"Invariant Object Recognition Based on a Neural Network of Cascade RCE Nets"* Proceedings of the IJCNN - 90, San Diego,Vol.2, pp. 845 - 854, 1990.

[211]- Hasegawa, A. & Shibata, K. & Itoh, K. &Ichioka, Y,Inamura, K. *"Adapting- Size Neural Network for Character Recognition on X-Ray Films"* Proceedings of the International Workshop on Application of Neural Networks to telecommunications, Alspector, J. & Goodman, R. & Brown, T. x. ,eds., Lawrence Elbaum Associates,Hillsdale,NJ,pp.139-146,1993.

[212]-Fukushima, K. & Miyake, S. & Ito, T. *"neocognitron: A Neural Network Model for a Mechanism f Visual Pattern Recognition"* IEEE transactions on Systems, Man, and Cybernetics, Vol. SMC -13, pp. 826-834, 1983.

[213]- Fukushima, K. *"Cognitron: A Self-Organizing Multilayered Neural Network"* Biological Cybernetics ,Vol. 20, pp. 121-136,1975.

[214]- Specht, D. F. "*probabilistic Neural Networks for Classifications, Mapping or Associative Memory*" Proceedings of the IEEE International Conference on Neural Networks, San Diego,Vol.1,pp.525-532,1988

[215]- Mood,A. M. & Graybill, F. A. "*Introduction to the Theory of statistics*" MacMillan, New York.

[216]- Parzen, E. "*On Estimation of a Probability Density Function and Mode*" Annalas of Mathematical Statistis,Vol.33, pp.1065-1076 ,1962.

[217]- Cacoullos, T. "*Estimation of a Multivariate Density*" Annals of Institude of statistical Mathematics,Vol.18, No.2, pp.179-189,1966.

[218]- Specht, D. F. "*A General Regression Neural Network*" IEEE Transactions on Neural Networks, Vol.2, No.6, pp.568-576, 1991.

[219]- Loskiewicz - Buczak, A. & Uhrig, R. E. "*Vibration Data Analysis Using probabilitic Neural Network-Based System*" Proceedings of INNS Meeting, World Congress on Neural Networks, Portland, OR, Lawrence Erlbaum Associates, Hillsdale, NJ, Vol.I, pp.273-278,1993.

[220]- Specht, D. F. "*Generation of Polynomial Discriminant Functions for Pattern Recognition*" IEEE Transactions on Electronic Computers, Vol.EC-16, pp .308-319, 1967.

[221]- Oja, E. "*A Simplified Neuron Model as a Principal Component Analyzer*" Journal of Mathematical Biology, Vol.15, pp.267-273, 1982.

[222]- Oja, E. "*Neural Network,Principal Components,and Subspaces*" International Journal of Neural Systems,Vol.1, pp.61-68, 1989.

[223]- Sanger, T. D. "*An Optimality Principle for Unsupervised Learning, in Advanced in Neural Information Proceeding Systems I*" Tourettzk, D. S., eds., Morgan Kaufmann, San Mateo, CA, pp.11-19, 1989.

[224]- Yuille, A. L. & Kammen, D. M. & Cohen. D. S. "*Quadratic and the Development of orientation Selective Cortical Cells by Hebb Rules*" Bological Cybernetics, Vol.16, pp.183-194, 1989.

[225]- Grossberg, S. "*Neural Expectation: Cerebellar and Retinal Analog Cells Fired by Learnable or Unlearned Pattern Classes*" Kybernetik, Vol.10, pp. 49- 57, 1972.

[226]- Von der Malsburg, C. "*Self-Organization of Orientation Sensitive Cells in the stirate Cortex*" Kybernetik, Vol.14, pp.85-100,1973.

[227]- Kohonnen, T. *"Self-Organization and Associative Memory"* 3rd ed., Berlin, Springer-verlag,1989-a.

[228]- Kohonen, T. *"A Self-learning Musical Grammer, or 'Associative Memory of the Second Kind"* International joint Conference on Neural Networks, Washington,DC, Vol.1, pp, 1-5, 1989-b.

[229]- Kohonen, T. *"Self-Organization and Associative Memory"* First ed., Springer- Verlag, Berlin-Heidelberg,1984.

[230]- DeSieno, D. *"Adding a Consience to Competitive Learning"* Proceedings of the IEEE International Conference on Neural Netorks, San Diego, Vol.1, pp.117-124,1988.

[231]- Kohonen, T. *"Improved Versions of Learning Vector Quantization"* International Joint Conference on Neural Networks,San Diego,CA,Vol.I, pp. 545- 550,1990-a.

[232]- Kohonen, T. *"The Self-Organizing Map"* Proceedings of the IEEE, Vol. 78(9), pp.1464-1480,1990-b.

[233]- Hecht - Nielson, R. *"Counterpropagation Network"* Applied Optics, Vol.26(23), pp.4979-4984,1987-a

[234]- Hecht-Nielson, R. *"Counterpropagation Networks"* IEEE First International Conference on Neural Networks, San Diego,CA,Vol.II, pp.19-32,1987-b.

[235]- Hecht-Nielson, R. *"Kolmogorov's Mapping Neural Network Existence Theorem"* IEEE First International Conference on Neural Networks, San Diego,CA,Vol.III, pp.11-14,1987-c.

[236]- Bogert, B. P. & healy, M. J. R. & Tukey, J. W. *"The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, Cross-Cepstrum, and Saphe cracking"* Proc.Symposium Time Series Analysis, Rosenblatt, M., ed. John Wiley and Sons, New York, pp.209-243,1963.

[237]- Carpenter, G. A. & Grossberg, S. *" A Massively Prallel Architecture for a Self-Organizing Neural Pattern recognition Machine"* Computer Vision, Graphics, and Image Processing, Vol.37, pp.54-115, 1987-a

[238]- Carpenter, G. A. & Grossberg, S. *"ART2: Self-organization of Stable Category* Recognition Codes for Analog Input Patterns" Applied Optics,Vol.26, pp.4919-4930,1987-b.

[239]- Lippmann, R. P. "*An Introduction to Computing with Neural Nets*" IEEE ASSP Magazine, Vol.4, pp.4-22, 1987.

[240]- Kalkunte, S. S. & Kumar, J. M. & Patnaik, L. M. "*A Neural Networks Approach for High Resolution Fault Diagnosis in Digital Circuits*" Proceedings of he IJCNN-92, Beijing,Vol.I, pp.I-83,I-88, 1992.

[241]- Smith, S. D. G. & Escobedo, R. & Caudekk, T. P. "*An Industrial Strength Neural Network Application*" Proceedings of INNS Meeting, World Congress on Neural Networks, Portland, OR, Lawrence Erlbaum Associate, Hillsdale, NJ, Vol. I, pp. 490 - 494, 1993.

[242]- Teow, L. N. & Lui, H. C. Wang, P. Z. & The, H. H. & Shen, Z. & Goh, T. H. "*Truth Value Flow Inference (TVFI) Neural Network*" private communication with [128]-Patterson, D. W.,1993.

[243]- Hsu, L. S. & The, H. H. & Chan, S. C. & Loe, K. F. "*Fuzzy Decision Making Based on Neural Logic Networks*" Proceedings of the Inter-Faculty Seminar on neuronet Computing, Technical Report DISCs No. Tra-6/89, National University of singapore, Singapore,1989.

[244]- mamdani, E. H. " *Application of Fuzzy Algorithm for Control of Simple Dynamic Plant*" Proceedings of the IEEE, Vol.121, pp. 1585-1589,1974.

[245]- Holland, J. L. "*Adaptation in Neural and Artificial Systems*" University of Michigan press, ANN Arbor, 1975.

[246]- Montana,D. & Davis, L. "*Training Feedforward Networks Using Genetic Algorithms*" Proceedings of the IJCAI-89, Vol.I, pp.762-767,1989.

[247]- Whitley, J. R. & Davis, J. F. "*Qualitative Interpretation of Sensor Patterns*" IEEE Expert, pp.54 - 63, April, 1993.

[248]- Mandischer, M. "*Representation and Evolution of Neural Networks*" IEEE Proceedings of the International Conference on Artificial neural Networks and Genetic Algorithm, Innsbruck, Springer-Verlag, Wien, pp.643-648,1993.

[249]- Mitchell, R. J. & Bishop, J. M. & Low, W. "*Using a genetic Algorithm to Find the Rules of a Neural Network*" IEEE proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithm, Innsbruck, Springer-Verlag, Wien, pp.664-669,1993.

[250]-Teh, A. H. & Tan, A. H. *"Connectionist Expert Systems-A Neural –Logic Model's Approach"* Proceedings of the Inter-Faculty Seminar on neuronet Computing, Technical Report DISCs No. Tra-6/89, National University of Singapore, Singapore, pp.16-32,1989.

[251]-Proakis, J. G. & Rader, C. M. & Ling, F. & Nikias, C. L. "*Advanced digital Signal Processing*" Macmillan publishing company, New York, USA, 1992.





السعر : 10 دولار أمريكي أومايعادلها